The Dissertation Committee for Justin Richard Garson certifies that this is the approved version of the following dissertation:

Psychiatric Disorders and Biological Dysfunctions: Some Philosophical Questions Concerning Psychiatry

Committee:

Sahotra Sarkar, Supervisor

Robert Causey

Robert J. Hankinson

John Z. Sadler

Richard Wilcox

Psychiatric Disorders and Biological Dysfunctions: Some Philosophical Questions Concerning Psychiatry

by

Justin Richard Garson, B. A.; M. A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2006

Dedication

This dissertation is dedicated to my father, John Garson, who struggled with mental illness for many years of his life, and who originally stimulated my interest in the subject of psychiatry, both through the example of his life as well as his own philosophical considerations on the nature of psychiatry. He initially disagreed with me on more or less everything that I originally stated here, but eventually I think we both came to moderate our viewpoints. (The quaint subtitle of the dissertation was his suggestion.)

Acknowledgements

My primary acknowledgement is first and foremost to Sahotra Sarkar, who took me under his wing at the outset of graduate school and taught me how to think, write, and present ideas as a philosopher and a scientist, over the course of seven years and several hundred pitchers of Texas' own Shiner Bock beer. I am grateful for his friendship and his mentorship. I also am deeply grateful to all of the other committee members: Robert Causey and Jim Hankinson of the philosophy department, who also played a large part in my philosophical training; Richard Wilcox of pharmacology, who played an important role in my understanding of neuroscience; and John Z. Sadler of psychiatry, whose philosophical articles were among the first to stimulate initially my interest in the link between mental disorder and biological dysfunction.

Psychiatric Disorders and Biological Dysfunctions: Some Philosophical Questions Concerning Psychiatry

Publication No._____

Justin Richard Garson, Ph.D.
The University of Texas at Austin, 2006

Supervisor: Sahotra Sarkar

Abstract: In contemporary biological approaches to psychiatry it is rarely questioned that psychiatric disorders stem from biological "dysfunctions". This assumption appears to be confirmed by the fact that biological research *has* been successful at uncovering diverse biological disparities between the brains of persons with mental illnesses and normal controls. However, the fact that something is different or unusual does not mean it is dysfunctional. The thesis of the dissertation is that there is little warrant for the claim that psychiatric disorders stem from biological dysfunctions. This prompts a question of definition: what does it mean to say that something – e.g., a given part of the brain or nervous system – is "functioning properly" or that it is "dysfunctional"? The dissertation argues that the theory of function appropriate for psychiatry is one that holds that the function of an entity consists in that activity that, in the past, contributed to the differential persistence or reproduction of that entity or type of entity. A consequence of this view is that just because something is not adaptive in a given environment, it is not necessarily dysfunctional. Finally, the dissertation examines

two major neurobiological perspectives on schizophrenia – a neurochemical perspective and a neurodevelopmental perspective. From a neurochemical perspective, it argues that even if the dopamine system is abnormal in schizophrenia, it is not dysfunctional. It also shows that on certain neurodevelopmental hypotheses, schizophrenia could be said to stem from a biological dysfunction, but on other neurodevelopmental hypotheses, it could not. The fact that there is currently not enough information to decide which of these multiple hypotheses is correct means that there is currently little warrant for saying that schizophrenia stems from a biological dysfunction Since this has been shown to be unwarranted through detailed analysis of some neurobiological examples, then it is reasonable to suspect that careful attention to neurobiological details associated with *other* mental disorders might reveal the same thing. Consequently, it should not be assumed that psychiatric disorders in general stem from biological dysfunctions on the part of the brain unless there is evidence for this conclusion other than the existence of biological abnormalities.

Table of Contents

List of Tables x

List of Figures xi

Chapter 1: Introduction 1

1.1 Thesis Overview 1

1.2 Relation of Thesis to Classic Debates: Antipsychiatry and the Medical Model 16

1.3 Relation of Thesis to Contemporary Debates: Descriptivism and Prescriptivism 29

1.4 Argument of the Dissertation 42

1.5 Chapter Overview 47

Chapter 2: From Mental Disorders to Internal Dysfunctions 49

2.1 Historical Motivation for Defining "Mental Disorder" 50

2.2 "Mental Disorder" in the DSM-III 74

2.3 Three Psychiatric Definitions of "Dysfunction" 90

Chapter 3: From Internal Dysfunctions to Etiological Functions 118

3.1 Taxonomy of Theories of Function 120

3.2 Uniqueness Claim for Etiological Theories 163

Chapter 4: From Etiological Functions to Selection Processes 186

4.1 Four types of Etiological Theory 188

4.2 Selection Processes in Psychology and Neurobiology 236

Chapter 5: Schizophrenia and the Dysfunctional Brain 269

5.1 Four Categories of Functioning 274

5.2 Neurobiological Approaches to Schizophrenia 286

Chapter 6: Conclusion: A Misbegotten Attempt 320

References.....326

Vita.....357

List of Tables

Table 2.1: Outline of positions on 1973 APA symposium on homosexuality 59
Table 3.1: Four types of etiological theory 129
Table 3.2: Addition of system and temporal variables to strong reproduction-based etiological theories 130
Table 3.3 Cases under which an entity can be dysfunctional 184

List of Figures

Figure 4.1: Four types of etiological theory 190
Figure 4.2: Example of complexity in the relation between X_1 and F_2 205
Figure 4.3: Example of complexity in the relation between F_2 and X_3 206
Figure 4.4: Four types of etiological theory 212
Figure 4.5: Four types of etiological theory 221
Figure 4.6: Innervation of skeletal muscle of newborn rats. 253
Figure 5.1: Functional motor neuron. 279
Figure 5.2: Motor neuron that is unable to function due to abnormal environment 281
Figure 5.3: Dysfunctional motor neuron. 282
Figure 5.4: Dopamine synapse 299

Chapter 1: Introduction

1.1 THESIS OVERVIEW

This dissertation begins with the *observation* that categories of mental disorder are *normative*; that is, in addition to describing how people actually behave, they describe how people should behave. More precisely, they describe how people actually behave in terms of deviation from a norm concerning how they should behave. This normativity prompts the following *question*: Where do these norms come from and how are they justified? According to one prominent view, these norms have a social origin and justification. In other words, the behavior, thoughts, or feelings of people with mental disorders deviate from social standards regarding appropriate conduct. According to another prominent view, these norms are biological in their origin and justification. In other words, mental disorders are symptoms of inner conditions, such as neurochemical imbalances or gross neuroanatomical variations, which deviate from standards of normal or proper biological functioning ("biological dysfunctions"). According to a third view, categories of mental disorder include both types of norms. It is also possible that the norms involved in mental disorder classification are psychological, ethical, or epistemological.

The *thesis* of the dissertation is that the norms appealed to in the context of psychiatric research and classification possess little biological justification. To express this thesis in a simple slogan, it is that *there is little warrant for the claim that psychiatric disorders stem from biological dysfunctions*. In other words, despite the continuous and substantial scientific and therapeutic advances in biological psychiatry, there is little warrant for believing that having a mental disorder involves a deviation from biological standards of proper functioning. This does not mean that mental disorders themselves do

not have biological bases, but that the *norms* that such categories express: "normal" versus "pathological", "functional" versus "dysfunctional", "appropriate" versus "inappropriate" – have little biological justification, nor is there reason to believe that such justification is forthcoming. This suggests that when psychiatrists implicate "dysfunctional" biological variation as a cause of a given mental disorder, the justification for the expression often involves an implicit appeal to the fact that the conduct of the person with the mental disorder deviates from external standards of appropriateness rather than intrinsic biological standards. The remainder of this section will expand upon this basic *observation* that motivates the dissertation, the *question* that it provokes, and the *thesis*.

The basic *observation* that motivates this dissertation is that the classification of a given psychological state or type of behavior as a mental disorder appears, on the surface, to involve a type of value-judgement that is very distinctive in the scientific context. Of course, the decision to accept any scientific theory is a value-laden one, in that it appeals to that theory's epistemic virtues, such as simplicity, explanatory or predictive power, or heuristic fruitfulness (Kuhn [1977, 331]). Medical concepts may be value-laden as well, in that the concept of "disease" or "pathology" seems to appeal to a norm, or set of norms, concerning the nature of individual well-being. Psychiatric classification, however, is distinctive in that the ascription of a given psychological or behavioral state to the status of a mental disorder often rests on judgements concerning the "inappropriateness" of the person's behavior, or psychological state, to a situation.

More specifically, a standard template for characterizing or defining a given mental disorder involves: a *person* who confronts a *situation*, where the situation *calls for* or *merits* a particular response (behavioral or psychological), and the response actually produced by the person is not an *appropriate* or *fitting* one. This response exceeds, falls

short of, or otherwise violates that which is appropriate to the situation. For example, according to the fourth, text-revised, edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR)[1], published by the American Psychiatric Association (APA) (APA [2000]) one of the diagnostic criteria for disorganized type schizophrenia is "flat or inappropriate affect" (315); for catatonic type schizophrenia, the "voluntary assumption of inappropriate or bizarre postures" (316); for major depressive episode, "feelings of worthlessness or excessive or inappropriate guilt" (356); for specific phobia, "marked and persistent fear that is excessive or unreasonable" (449); for schizotypal personality disorder, "odd, eccentric, or peculiar" behavior and appearance, as well as "inappropriate or constricted affect" (701); for antisocial personality disorder, "failure to conform to social norms with respect to lawful behaviors" in addition to "consistent irresponsibility as indicated by repeated failure to sustain consistent work behavior or honor financial obligations" (706). One who has obsessive compulsive personality disorder may be "excessively devoted to work and productivity to the exclusion of leisure activities and friendships", and may be "overconscientious, scrupulous, and inflexible about matters of morality, ethics, or values" (729); one with histrionic personality disorder may exhibit "inappropriate sexually seductive or provocative behavior" as well as an "exaggerated expression of emotion" (714); the manic episode is characterized by "excessive involvement in pleasurable activities that have a high potential for painful

[1] Throughout this dissertation, "DSM" refers to some edition of the *Diagnostic and Statistical Manual of Mental Disorders*, which is the American Psychiatric Association's official classification of mental disorders. There are currently four editions that differ in important ways, hence they will be identified with the appropriate edition number: DSM-I (APA [1952]); DSM-II (APA [1968]); DSM-III (APA [1980]); and DSM-IV (APA [1994]). Additionally, there is a revised version of the DSM-III, labeled DSM-III-R (APA [1987]), and the most current edition, DSM-IV, also has a text revision, labeled DSM-IV-TR (APA [2000]). Finally, "APA" will replace all future references to "American Psychiatric Association", but since the American Psychological Association has the same initials, the name of the latter organization will always be rendered in full.

consequences" such as "engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments" (362).

The basic question that this observation provokes, then, is the following: What are the *origin* of and *justification* for the standards or norms of appropriateness that are embodied in mental disorder ascriptions?

There are two very general views that, by and large, have structured the debates concerning the origin and justification of these standards, each of which has strong implications for the self-conception of psychiatry, its domain of investigation, and its relation to other disciplines. The first position will be referred to as the "social values" position; the second the "biological dysfunction" position. These two positions are not exhaustive, but they are the most prominent.

With respect to the *origin* of these standards of appropriateness, the first position holds that these standards originate from widely-held social values. Each society constructs a code of mores or expectations that are applied to each person within it; mental disorder categories simply define rather extreme ways in which people can deviate from this code.[2] For example, with respect to behavior, it is not socially *appropriate* to speak rudely to superiors, to laugh during funerals, or to undress in public. Such standards are not only applied to public behavior, but also to "private", psychological states, such as the emotions that motivate one's behavior or the rationale that justifies one's behavior. Certain situations are widely held to be pleasurable:

[2] E.g., Szasz (1961); Scheff (1966); Sarbin (1969). This view was prominent in the so-called "antipsychiatric" tradition; see Section 1.3. However, this view is by no means restricted to that tradition, and is expressed in some conventional psychiatric textbooks. For example, Redlich and Freedman (1966) define "behavior disorder" as "behavioral patterns…that are not compatible with the norms and expectations of the patient's social and cultural system (quoted in Boorse [1982, 38])"; Ullmann and Krasner (1966) define "maladaptive behavior" as "behavior that is considered inappropriate by those key people in a person's life who control reinforcers (quoted in Boorse [1982, 40])". Ausubel (1961) defines "mental illness" in terms of "gross deviation from a designated range of desirable behavioral variability (Ibid., 72)", although he does not specify the agent or agents whose desires are relevant to determining the appropriate such range of variation.

advancing in one's chosen career, meeting with good friends over dinner, or pursuing a long-term intimate relationship. Someone who displays a lack of interest in these pleasures, for example, who displays indifference toward praise or criticism,[3] or exhibits emotional withdrawal from friends or family,[4] or who derives sexual gratification from unusual objects or situations,[5] seems not to recognize the pleasurable events in life, and therefore from the point of view of these social standards, the person's emotions reflect an *incorrect evaluation* of the value of the situation. Similarly, some explanations for one's behavior are thought to be *reasonable* ("I rushed into the burning building to save my friend"), while some are *unreasonable* ("I rushed into the burning building to save my goldfish"), and some are simply incoherent or *bizarre* ("I rushed into the burning building because God told me to"). To say that a belief is "bizarre" is to say that it deviates widely from the expectations and beliefs that are widely shared within a community.[6]

From the "social values" position, then, the standards of appropriateness embodied in psychiatric concepts, whether they refer to behavior, thought, or feeling, describe a lack of fit between the behaviors, thoughts, or feelings of the individual and the values held by the society at large. This does not mean that the individual does not share the same values held by the society at large, or does not recognize his or her own actions or attitudes to be in conflict with them, or does not wish to seek the help of mental health professionals in order to change them. For example, a criterion of specific phobia is that one acknowledges his or her fear to be excessive or unreasonable (APA [2000, 449]).

---

[3] E.g., schizoid personality disorder (APA [2000, 697]).
[4] E.g., posttraumatic stress disorder (APA [2000, 468]).
[5] This is a defining feature of the paraphilias (APA [2000, 536]).
[6] The standard psychiatric definition of a "bizarre delusion" is "a false belief that involves a phenomenon that the person's culture would regard as totally implausible" (APA [1987, 395]).

The question of the *justification* of such standards – that is, the question of why *these* standards, rather than others, should be brought to bear on the individual – can often be answered by an appeal to pragmatic social grounds. At one extreme, a person diagnosed as having a mental disorder may be unpredictable, in that he or she acts in ways that seem bizarre or unruly, and therefore may represent a danger or threat to others. In this case, the revocation of certain liberties, for example, the involuntary hospitalization of the person so diagnosed, may be warranted for the sake of protecting the other members of society. At a lesser extreme, the person so diagnosed may pose a risk to himself or herself. The condition in question may result in job loss, social exclusion or stigmatization, or, if the disorder has a strong affective component, the possibility of suicide. Hence it may be beneficial to the person so diagnosed if treatment plans are made available to that person, if medical insurance providers are willing to cover for such treatment plans, and if the person has some amount of legal protection with respect to potentially discriminatory employment practices. The question of what *degree* of potential harm a person must represent to others, or to self, in order to justify the revocation of liberties or the bestowal of privileges, is a nuanced ethical, economic, and political question that will not be broached here (Robinson [2003]; Edwards [1982]). However, from this perspective, there is no principled reason why, e.g., racism should not be judged to be a mental disorder,[7] along with schizophrenia, bipolar disorder, and anti-social personality disorder: extreme racism represents a form of conduct or belief that violates widely-shared standards of appropriateness and that is associated with harmful behavior.

[7] See Poussaint's op-ed in the New York Times ("They Hate. They Kill. Are they Insane?"), which recommends that extreme racism be recognized as a major psychiatric illness (26 August 1999), as well as a number of dismissive rejoinders that shortly followed ("Racism is Not a Treatable Illness", 30 August 1999; "Classifying Racism as Insanity isn't that Easy", 31 August 1999).

However, the social values position also leaves open the possibility that there are behavioral or psychological conditions that are *currently* recognized as mental disorders according to standard classification and diagnostic systems, but that are relatively harmless and that largely reflect overly narrow or parochial social judgements. In this case, on the basis of the principles that govern a liberal society, it may be preferable if those social standards were themselves modified or expanded to embrace, or at least tolerate, these unusual conditions. For example, the delisting of the category of homosexuality from official APA nomenclature in 1973 probably reflects, in part, a shift in widely-held social values with respect to sexual orientation that occurred during this period.[8]

From this perspective, then, the domain of investigation of psychiatry amounts to the explanation, prediction, and correction of deviance, in thought or action, from standards of appropriateness that originate from within, and possess whatever justification they have by virtue of, a person's social sphere. In this respect, the domain of psychiatry borders upon that of sociology, on the one hand (which explicates and enumerates such social codes), and that of law, on the other (which seeks to correct such violations). This position does not imply, however, that biological medical intervention constitutes an inappropriate treatment method. It does, however, imply that such intervention (e.g., pharmacological intervention for schizophrenia) has the same status as biological intervention in the case of, e.g., sexual offenders who are not typically deemed to have mental disorders. In both cases, a socially undesirable and potentially harmful disposition is alleviated or corrected, ideally for the benefit of all parties.

---

[8] To say that homosexuality was "delisted" at this time is actually an overstatement. After a series of debates within the psychiatric community in the early 1970s (see Stoller *et al*. [1973]) the category was renamed "Ego-dystonic Homosexuality" and appeared as such in DSM-III (APA [1980, 281]). The category was eventually dropped only in 1987 with the publication of DSM-III-R (APA [1987]). See Section 2.1.1 for a more extensive overview of the history of this debate.

The second major view concerning the origin and justification of the standards of appropriateness embodied in mental disorder categories, the "biological dysfunction" position, is one that is more consistent with the current thinking and practice of most biologically oriented psychiatrists. With respect to the *origin* of such standards, they are held to stem from the biological sphere itself, rather than the social sphere. As is widely believed throughout several biological and medical disciplines, the various parts of the organism, and perhaps the organism itself, are subject to standards of *functioning*. In healthy individuals, the biological parts or activities of the organism function *normally* or *properly*; in the case of disease, some organ or activity is said to be *malfunctioning* or *dysfunctional*. To say of a given biological entity that it is *functioning properly*, or that it is *malfunctioning*, is to appeal to standards or norms of proper biological functioning that the entity in question may or may not satisfy. Hence the concept of a biological function does not necessarily *describe* the current activity of a given entity, but sets up a standard or norm by which the activity of the entity is evaluated. In this sense, standards of biological functioning are normative. To use a paradigmatic example, the function of the eye is to see. This does not describe what all eyes do; rather, it is often used to articulate the intuition that seeing is what eyes are "supposed to" do, what they are "there for", or what they have the "purpose" of doing. If they do not enable a person to see, it is sometimes said that they are "malfunctioning" or "dysfunctional". This provokes the question as to how such biological standards originate, and whether they are in fact explicable purely on the basis of biological considerations.[9]

[9] In the following, there is no intended implication that the "social realm" and the "biological realm" are mutually exclusive. In Section 2.3.2, it will be argued that whether or not a given condition ultimately has a "biological" or "social" cause is irrelevant to whether or not it currently stems from a "biological dysfunction", and hence the importance of the debate about the relative role of biological or social causes in the etiology of mental disorders will be minimized from the standpoint of the dissertation.

Depending on the concept of "biological function" that is appealed to, these standards or norms may have one of several distinct biological origins. They may be evolutionary in origin: perhaps an organ is functioning *properly* when it is performing the activity that it was selected for by natural selection to perform, and it is *malfunctioning* otherwise. Alternatively, such standards may be related to the current adaptiveness of the performance of an activity, regardless of its evolutionary origin: perhaps a part is functioning properly when it contributes to the survival or reproduction of the organism which contains it, and malfunctioning otherwise. A third alternative – one which restricts the concept of functioning to sentient creatures – is that these standards of functioning are partly psychological in nature; a part is functioning properly when it contributes to the well-being of the organism that possesses it, that is, to a relative freedom from physical suffering or to the organism's capacity to pursue self-chosen goals, and "dysfunctional" otherwise, regardless of whether the dysfunction is reproductively disadvantageous for the organism or has been selected against historically. This last alternative, of course, does not reduce the question of the origin of standards of functioning to purely biological considerations, but, in addition to biological considerations, it involves some prior conception about the nature of personal well-being for sentient creatures.[10]

According to this perspective, mental disorders are, or stem from, biological dysfunctions on the part of the individual. They are differentiated from physical disorders in that they typically undermine proper psychological functioning rather than proper

[10] Sedgwick (1981), for example, holds that the concept of disease generally, whether mental or physical, rests on a value-judgement concerning the undesirability of an otherwise value-neutral biological condition (see Section 1.3 for an elaboration of this position, according to which "disease" is an *evaluative term* that does not possess much in the way of specific descriptive content). However, the diagnosis of tuberculosis or cancer does not necessarily appeal to standards of the "appropriateness" of the afflicted person's behavior or psychological state with respect to the social sphere, and therefore remains less problematic than the ascription of a mental disorder.

physiological functioning.[11] They may inhibit the person's capacity to engage in means-ends reasoning, or to produce an accurate internal representation of his or her current situation, or they may lead to despair or suffering on the part of the individual. They may also eventuate in the person's inability to comprehend or conform to widely-held social values, but nonetheless, the inappropriate behavior that the person exhibits is not *constitutive* of having a mental disorder, but one of its *symptoms* or *effects*. (Consequently, deviation from a social code of conduct may be legitimately used as a heuristic for the identification of a mental disorder.) What transforms a given biological *cause* into a biological *dysfunction* is merely that it falls short of biological standards of functioning. In this way, schizophrenia is analogous to cancer, heart disease, or diabetes, and does not involve presuppositions that are any more problematic than those involved in determining that the latter conditions constitute diseases.[12] From this perspective, racism is very unlikely to constitute a mental disorder, since it is probably not a symptom of a biological dysfunction.

The *justification* for applying these biological standards or norms of well-functioning to the organism or to its parts stems merely from the recognition that humans are biological in nature and therefore that biological activity on the part of humans can legitimately be judged in terms of the norms that govern the remainder of the biological sphere. However, the justification for medical intervention will rely on some of the same considerations discussed earlier, e.g., the recognition of the potential harm or suffering caused by the dysfunctional item.

---

[11] In Chapter 2 it will be argued that the most consistent development of the "biological dysfunction" position should hold that the concept of a mental disorder has no substantive differentia over and above non-mental medical disorders, but that this difference is purely methodological or heuristic.

[12] To make the proposed analogy more concrete, one might say that cancer results from the overproduction of infected cells, just as schizophrenia may result from the overproduction of dopamine.

From this perspective, the goals of psychiatry are more intrinsically aligned with those of medicine rather than law, insofar as it is concerned with the treatment of biological dysfunctions on the part of the individual rather than the elimination or modification of socially undesirable conduct, and insofar as it utilizes the methods of the natural sciences, such as those found in neuroscience, molecular biology, and evolutionary biology, to understand the etiology of those dysfunctions, and on that basis, to devise rational treatment plans or to predict treatment outcomes.

This is not to say that social norms or values, according to this perspective, are not *relevant* to the classification of a given type of psychological state or behavior as a mental disorder. For each proposed mental disorder category, it must be asked whether having the mental disorder represents a sufficiently grave threat to society, or to the person so diagnosed, to warrant the ethical, legal, social, and economic consequences of its subsumption within an official diagnostic or classification scheme.[13] Nonetheless, the appeal to social values within the context of psychiatric classification and research would be secondary to the determination that the aberrant behavior or thinking in question stems from a biological dysfunction on the part of the individual.[14]

---

[13] For example, transitory cognitive impairment (TCI) is often considered to be a "sub-clinical" type of epilepsy, in that it is not associated with pronounced seizures (although it manifests itself in performance on certain specific cognitive tests). To the extent that it represents a mild form of epilepsy it is thought to stem from a biological dysfunction. However, it is debatable whether the symptoms of TCI result in a sufficient degree of social maladaptiveness or harm to be included within a standardized medical or psychiatric nomenclature (Binnie [2003]). Hence, from the "biological dysfunction" position, there is an important distinction to be drawn between the sort of justification involved in the professional authorization and application of the "mental disorder" label to a given condition, and the sort of justification involved in determining whether someone *has* a mental disorder in the first place. See, e.g., Sarbin (1967) and Ellis (1967), who debate the merits of the institution of labeling on the basis of its social, ethical, and therapeutic consequences.

[14] As noted above, there also exists a third, prominent position, which is a combination of the "social values" position and the "biological dysfunction" position. According to this view, mental disorders involve violations of *both* types of normativity: they are caused by biological dysfunctions on the part of the individual and they also violate social standards of appropriate conduct. However, this position is relatively unimportant from the perspective of the dissertation, which is more exclusively concerned with analyzing the appropriateness of appealing to biological norms at all in the psychiatric context.

Having described the *observation* that motivates the dissertation (that mental disorder concepts are laden with norms of appropriateness), and the basic *question* that the dissertation seeks to resolve (the origin and justification of these norms), the *thesis* of the dissertation can be stated. The thesis of the dissertation is that *appeal to the notion of a "biological dysfunction" in the context of psychiatric research and classification often presupposes standards of appropriateness which have no biological justification*. The problem is that it is often assumed that just because, e.g., the brain of a person with schizophrenia is different from the brain of a person without it, then the schizophrenic brain must be "dysfunctional". But just because something is different does not mean that it is dysfunctional! Often, in the context of psychiatric research, the judgement that something within an individual is specifically "dysfunctional" (rather than merely "different" or "uncommon") is based solely upon the fact that its behavioral or psychological effects are inappropriate to various situations that the individual confronts in everyday life. Therefore, far from reducing these standards of appropriateness to biological standards of functioning, the notion of a biological dysfunction as it is often used in psychiatry presupposes them.

This conclusion does not imply that these standards or norms of appropriateness are social in their origin and justification. They may also appeal to psychological norms, or to universal ethical, epistemic, or aesthetic values. For example, as noted above, a criterion for antisocial personality disorder is "consistent irresponsibility as indicated by repeated failure to sustain consistent work behavior or honor financial obligations" (APA [2000, 706]). This appears to represent an *ethical* norm. The purpose of the dissertation is not to engage in an analysis or evaluation of the type of norms involved in psychiatric research and classification (e.g., Sadler [2004]), but merely to reject a prominent viewpoint about the nature of those norms.

The fact that biological dysfunction ascriptions made in the context of psychiatry often presuppose norms of appropriateness the origin and justification of which cannot be biologically established does not undermine the scientific status of psychiatry as a discipline. Many disciplines are inherently normative in nature, where these norms have their origin and justification on the basis of social or ethical values. Conservation biology (which is concerned with methods for successfully conserving biodiversity, where the conservation of biodiversity is an important social, ethical, and aesthetic value – see Soulé [1985]) is a paradigmatic example of such a discipline. However, it does suggest that the appeal to these values in the context of psychiatric classification and research should be rendered as explicitly as possible, and not concealed by biological terminology – such as "dysfunctional", "dysregulated", "chemical imbalance", "neurohormonal disturbance", etc. – that appear to attribute these norms to the biological sphere in ways that are not warranted by biological research.

This does not mean that such terminology is never warranted or applicable in the medical context, or even that it is impossible for a psychiatric disorder to be shown to originate from a biological dysfunction, but that typically, in fact, it has little warrant in the psychiatric context, and there is little reason to suggest that such warrant is forthcoming. Three caveats, then, about the epistemological and ontological status of the thesis are in order:

> (i) the thesis that appeal to the notion of a "biological dysfunction" in the context of psychiatric research is not biologically justified does not bear conceptual necessity. It is not a purely conceptual claim; rather, it is entailed by conceptual as well as empirical premises. It involves conceptual considerations insofar as it presupposes the proper explication of the concept of a "biological dysfunction"

that is appropriate to psychiatry. It involves empirical considerations in that the empirical evidence currently available concerning the biological bases of psychiatric conditions typically does not warrant the claim that a given psychiatric condition stems from a biological *dysfunction* (although it may stem from a biological *cause*);

(ii) the thesis does not have the status of a universal generalization about biological (dys-)function ascriptions applied to mental disorders. Epilepsy, Huntington's disease, and general paresis, are examples of conditions that can probably correctly be said to stem from biological dysfunctions on the part of the individual, and which can eventuate in delusional, disoriented, or demented psychological states, as well as to behaviors that are socially inappropriate. Thus, consistent with the "biological dysfunction" position, such psychiatric manifestations should be counted as "symptoms" of an underlying dysfunctional process (e.g., synchronous waves of electrical discharge throughout large areas of the brain [epilepsy]; late stage syphilis [general paresis], and trinucleotide repeat disorder [Huntington's disease]). However, with respect to one major category of mental disorder that has gained substantial contemporary research attention by biologically-oriented psychiatrists – schizophrenia – the current understanding of the biological correlates of this condition does not warrant the claim that schizophrenia stems from a "biological dysfunction". As will be discussed below, the rationale employed to arrive at this conclusion should be generalizable to other mental disorders as well; and

(iii) the thesis, however, is not simply a reiteration of the fact that there is much that is biologically unknown about the etiology of most mental disorders – a fact that is routinely acknowledged by psychiatrists as an incentive for further research. But in acknowledging this, there is typically little or no consideration about what *sort* of etiological account of the biological basis of a mental disorder would qualify as showing that the psychiatric symptoms result from a biological *dysfunction* (rather than, e.g., "normal variation" in a biological trait). It is important, therefore, not just to provide an overview of current evidence and research, but to go beyond the given evidence to examine broader theoretical perspectives on mental disorders, and evaluate the extent to which the following question can be answered in the affirmative: *Were* this theory to be well-confirmed in the future, would one be warranted in claiming that this mental disorder stems from a biological dysfunction on the part of the individual? For example, according to the "sensory-gating" theory of schizophrenia (e.g., Grace [2000]), the so-called "positive symptoms" of schizophrenia (hallucinations and delusions) are a consequence of the relaxation of sensory filtering mechanisms that lead to sensory bombardment. Although there is some evidence for this theory,[15] all that is strictly speaking derivable from the theory is that there is variation in the strength of sensory gating across schizophrenic and non-schizophrenic populations; it does not alone explain why a certain degree of variation should be conceived of as dysfunctional or pathological. The latter judgement is a further theoretical and empirical claim that involves for its justification (according to this dissertation) the determination of the "proper function" of the mechanism, the range of environments under which, historically,

[15] See Heinrichs (2001, 74-75) on the P50 evoked potential "gating" defect as a reliable marker for schizophrenia.

the mechanism came to possess the function, and the range of activity which is consistent with performance of that function. Once such information is available, then the variation in question *may* be shown to qualify as "dysfunctional" variation. Nonetheless, many other possibilities may emerge, for example, that the biological variation in question is within its historically "normal" range, or, that though the biological variation falls outside of this normal range, it does so because it is placed within a historically abnormal context and therefore it is not necessarily inherently dysfunctional.

The remainder of this introductory chapter will accomplish the following three objectives. First, it will place the thesis of the dissertation in the context of classic debates on the value-ladenness of psychiatric classification ("Antipsychiatry and the Medical Model"; Section 1.2), as well as in the context of current debates on the value-ladenness of psychiatric classification ("Descriptivism and Prescriptivism"; Section 1.3). In doing so, it will clarify the thesis by providing a contrast between the framework presented here and those of the classic and contemporary debates. Secondly, a schematic overview of the argument that will be presented in the dissertation will be provided (Section 1.4). Thirdly, a section overview will be provided (Section 1.5).

1.2 Relation of Thesis to Classic Debates: Antipsychiatry and the Medical Model

To clarify the thesis of the dissertation – that there is little warrant for the claim that psychiatric disorders stem from biological dysfunctions – it is helpful to contrast this thesis with one that was popular within the antipsychiatric tradition that flourished in the 1960s and early 1970s, particularly in France, Britain, and the US.[16] In the following,

[16] E.g., Szasz (1961); Goffman (1961); Laing and Esterson (1964); Scheff (1966); Foucault (1967); Cooper (1967); Sarbin (1969).

after providing an informal overview of the "antipsychiatry" tradition and its relation to the "medical model" of psychiatry, some of the central assumptions governing both traditions will be more formally identified and criticized. The thesis of the dissertation will be shown to be *compatible* with a weakened version of the antipsychiatric empirical-sociological claim that the purpose of psychiatry as an institution is the explanation, prediction, and correction of certain forms of deviation from norms of appropriateness that are social in their origin and justification. However, it will be pointed out that the formulation of this thesis does not necessarily imply the stronger normative claims about treatment methodology that the antipsychiatrists advocated; for example, it does not suggest that biomedical treatment of mental disorders is inappropriate and should be replaced by community-based treatment, or that "role-playing" models of mental disorder are valid.

1.2.1 Overview and Criticism of Antipsychiatry

One of the predominant themes in the antipsychiatric tradition is the notion that the behavioral and psychological patterns typically conceived of as symptoms of mental disorders (e.g., delusions, hallucinations, etc.) cannot be conceptualized in abstraction from the complex set of social interactions within which a person comes to adopt a certain social role (namely, that of the "mentally ill person", or the "sick role"[17]). The reasons that a person adopts this role, moreover, are typically held to be explicable in terms of normal interpersonal situations or stressors, and hence do not need to be explained in terms of a peculiar or "special" inner cause, whether psychological or biological. Consequently, proponents of antipsychiatry were equally opposed to

[17] See Parsons (1951) for the theoretical framework that defines the "sick role"; this notion will be elaborated in Section 2.3.3.

biological as well as psychodynamic psychiatry, insofar as both tend to be exclusively individualistic.

For example, Cooper (1967), who coined the term "anti-psychiatry", defines schizophrenia in the following terms:

> Schizophrenia is a micro-social crisis situation in which the acts and experience of a certain person are invalidated by others for certain intelligible cultural and micro-cultural (usually familial) reasons, to the point where he is elected and identified as being "mentally ill" in a certain way, and is then confirmed (by a specifiable but highly arbitrary labeling process) in the identity "schizophrenic patient" by medical or quasi medical agents. (Ibid., 2)

The sociologist, Scheff (1966), similarly conceives of mental illness as the outcome of generic social reinforcement techniques by which normal deviation from social standards is aggravated, and implicitly rewarded, by the community which evaluates it as such. (Similar views are expressed in Szasz [1961], Goffman [1961, especially pp. 350-366]; Laing and Esterson [1964], Sarbin [1969], and Rosenhan [1973].) The reason that the label "antipsychiatry" is often used to describe these views is that insofar as psychiatric institutions perform the role of the "medical or quasi medical agents" described by Cooper which "confirm" the mental illness label, they are thereby thought to perpetuate, by reinforcement, the very conditions that they are assigned to ameliorate.

It is important to point out that many of the conditions deemed by more conventional psychiatrists to constitute mental disorders are also conceived of by the antipsychiatrists as forms of human suffering that are prima facie undesirable to have. Consequently, the label "antipsychiatry" is perhaps unwarranted, since many of those who fall under the category (such as Szasz, Cooper, and Laing) are themselves psychiatrists who recognized a need for help and advocated more interpersonal or

community-based modes of treatment. One of the motivations that fueled the antipsychiatry approach, then, was their view that they were in possession of better, more effective techniques of ameliorating human suffering. To this extent, "antipsychiatry" (or at least the version presented here) refers to a set of overlapping theoretical frameworks from which empirically testable hypotheses concerning the alleviation of certain forms of suffering can be derived.[18]

Having provided an informal overview of the tradition, some of the central assumptions of that tradition can be more formally identified and criticized. Despite variations in their formulations, advocates of the antipsychiatric perspective typically coupled two claims, one empirical-sociological and the other conceptual.[19] The empirical claim is that psychiatry is an institution that has the function of regulating social deviance. The conceptual claim is that the statement that "Person *P* has a mental disorder, *D*" *either*:

> (i) does not attribute a non-relational property to an individual. Rather, it describes a relationship between the psychological or behavioral state of an individual and the values held by a *particular* social community. In other words, to say that a person has a mental disorder is to say that the person's behavior is deviant with respect to that particular society's mores or values; *or*

[18] For example, much of the empirical impetus that motivated the social-role theory of schizophrenia stemmed from work in the US by Bateson and colleagues on the "double-bind" theory of schizophrenia (Bateson *et al*. [1956]), according to which schizophrenic symptoms such as delusions and hallucinations, catatonia, and disorganized thought can be understood as intelligible responses to conflicting demands that are imposed upon a person (typically by the family, and most commonly by the mother) each of which is associated with punishment and which jointly admit of no satisfactory resolution. This work formed the theoretical basis for Laing and Esterson (1964) and Cooper's (1967) resulting "definition" of schizophrenia (above).

[19] This rendering of the two theses simplifies the antipsychiatric approach and excludes reference to the antipsychiatrists' moral and political contention that medical psychiatry is inhumane or that it produces undue suffering. See Dain (1994) and Wilson (1993) for more detailed and scholarly overviews of the moral and political aspects of antipsychiatry.

(ii) the statement does not *describe* anything at all, but is purely *prescriptive*. In other words, to say that a person has a mental disorder is to say that the person ought to be excluded from normal social intercourse, or that the speaker, or the institution of which the speaker is a representative, does not like the person's behavior, etc. "Mental disorder", according to this view, has the status of an epithet for expressing a negative attitude about someone's conduct. In this respect it is similar to ethical emotivism (Stevenson [1937, 18]; Ayer [1952, 107]), according to which the function of moral terms is to evince disapproval of something and to incite similar emotional attitudes in others.

Although from a philosophical point of view, version (i) and version (ii) of the conceptual claim have very different implications – one endorses relativism and the other non-cognitivism about "mental disorder" ascriptions – both versions lead to the same critical conclusion that, insofar as the values that are expressed in prescriptive norms often change as a function of social context, the sorts of behavioral or psychological conditions that may warrant or provoke the "mental disorder" label in one society or era will not necessarily warrant or provoke the "mental disorder" label in another society or era.

Often the conceptual claim is supported by providing a social history of a given mental disorder category and revealing correlations between changes in the psychiatric status of the condition, on the one hand, and changes in widespread social values and economic structures, on the other. This correlation is supposed to show that changes in the public conception of what qualifies as a mental illness is best explained as a function

of widespread social values and mores, rather than medical or scientific facts.[20] The conceptual claim is also often supported by providing cross-cultural evidence that patterns of behavior that in one society constitute evidence for a mental disorder may, in another society, be considered "normal" or even commendable. A well-worn example, taken from Silverman (1967), is of the trance-like behavior characteristic of the Siberian shaman, which would be considered evidence for schizophrenia in the West, although the shaman performs a valuable and respectable social role in his own social context.[21]

The empirical claim is typically supported by documenting the history of the way in which psychiatric institutions largely succeeded in excluding those labeled mentally ill (who were often "social deviants", such as unrepentant alcoholics, criminals, or vagrants) from normal human intercourse: through involuntary hospitalization, abusive medical intervention, or merely by the social stigmatization that invariably followed a diagnosis.[22] Such documentation is used to suggest that the true function of psychiatry is to regulate social deviance and that any claim to the contrary is ideological.

Both theses are unnecessarily strong, and neither will be advocated here. With respect to the empirical claim, if the expression "social deviance" is taken in its common sense, e.g., "criminal", "unruly", or "disruptive of the social order", then it is false. Depression, anxiety, or Alzheimer's dementia, for example, are not "socially deviant" in any of these senses, yet they are all widely considered to be mental disorders. Similarly,

---

[20] This is presumably why Foucault (1967) is often classified as "antipsychiatric", although Foucault does not in that context explicitly expound any general antipsychiatric position.

[21] See, however, Murphy (1978) for a critical evaluation of this claim from an anthropological perspective. In her own ethnographic work with this group she finds no overlap between those considered insane (*nuthkavihak*) and those considered shamans; that though shamans exhibit some of the behavior of the *nuthkavihak*, they do so exclusively in the context of circumscribed, ceremonial occasions; that those considered *nuthkavihak* do not typically occupy useful social roles; and that most of the characteristic behaviors of the *nuthkavihak* would also be considered signs of mental disorder in the West (Ibid., 6).

[22] For a recent text, see Whitaker (2002); however, this sort of polemical historiography of psychiatry is currently undergoing a shift towards a more balanced assessment of the historical role and function of psychiatric institutions (see Porter and Wright [2003] for a recent anthology).

criminal violence or anti-war protest may be considered to be "socially deviant", but they are not necessarily considered to be mental disorders.[23] (In the common sense of "social deviance", antisocial personality disorder and conduct disorder would be two of the only psychiatric conditions that are largely defined in terms of a history of unruly, disruptive, or violent behavior). However, the thesis of the dissertation is compatible with a weaker sociological claim: Psychiatry is an institution one of the functions of which is to explain, predict, and correct certain forms of deviation from norms of appropriateness that are social in their origin and justification. This does not imply that such norms are illegitimate or unjustified, or that the resulting applications of social and political power are wrong.

The conceptual claim, according to which mental disorder concepts are largely relative to the values of a particular society, appears to import a thesis about social or historical relativism into the view that mental disorder categories embody social values: namely, that if a statement is laden with social values then the statement is held to be true in some societies and not in others. But a statement may be at once laden with social values, and nonetheless cross-culturally or cross-generationally shared. For example, it may be a universally shared moral platitude that taking personal advantage of communal resources is a disreputable thing. Thus, even if there exists anthropological evidence that certain types of behaviors are universally considered to be signs of mental disorders, this would *not* uphold the thesis that categories of mental disorders are not laden with social values of appropriateness.[24]

---

[23] Boorse (1982, 39); Wakefield (1992a).

[24] Similarly, even if there were biological evidence that some social values, such as the prohibition of taking personal advantage of communal resources, are themselves a product of evolution, this does not entail that the person who violates such norms suffers from a biological dysfunction. The thesis endorsed here does not even imply the conceptual possibility that a society could exist that does not hold certain norms. For example, it may be conceptually impossible for there to be a society that did not impose a prohibition on taking personal advantage of communal resources.

Similarly, the converse of the above implication – that if a statement is held to be true in some societies and not in others, then the statement is laden with social values – is also invalid. The proposition that epilepsy is a disease is not universally recognized as such, but that does not *entail* that its endorsement is laden with social values. For example, among the Hmong in Laos, epileptic seizures are commonly thought to indicate divine blessings (Fadiman [1997]).[25] However, this does not mean that epilepsy is not "really" a disease, or that in calling epilepsy a "disease" one is tacitly expressing a set of socially-relative value judgements. It may suggest that being a "disease" and being a "blessing" are not incompatible;[26] it may also suggest that that the nature of epilepsy is widely misunderstood in some cultures.[27]

Although the failure of this inference is fairly obvious, it has an important implication that has been largely unrecognized by advocates of the antipsychiatric position, which is that anthropological or historical evidence that certain mental disorder ascriptions are not, or have not been, universally recognized as such, or that they change over time, does not imply that they do not really stem from biological dysfunctions on the part of the individual. Consequently, although the thesis of the dissertation is that biological dysfunction ascriptions made in the context of psychiatry are typically biologically unwarranted, evidence for cultural variation in the application of mental disorder concepts will not be invoked to establish this thesis.

---

[25] The association between epilepsy and divine causation has a long history in the West; Hippocrates (1952; see "On the Sacred Disease"), for example, explicitly denounces the view that epilepsy is a "sacred disease", arguing that it is caused by the brain and should be treated like other diseases, rather than by religious ceremonies.

[26] According to Fadiman (1997, 20) epilepsy possesses this dual status among the Hmong.

[27] Neander (1983, 30).

1.2.2 Overview and Criticism of the "Medical Model" of Psychiatry

Advocates of the antipsychiatric position typically pitted themselves against proponents of the so-called "medical model" of psychiatry, who commonly endorse opposing conceptual and empirical claims.[28] They reject the empirical claim that the function of psychiatry is to regulate social deviance. Instead, they hold that psychiatry is a branch of medicine, and its function is to treat diseases. The conceptual claim is that to say that "Person *P* has a mental disorder, *D*" is to attribute a non-relational property to an individual. Often, the "medical model" in psychiatry is associated specifically with the idea that this property is a biological property, such as a disorder of the brain or nervous system. However, at the time of these disputes, many psychodynamic psychiatrists thought of themselves as practicing a form of psychological medicine – psychoanalysis – that isolates and cures unconscious psychological conflict in a manner analogous to the way in which physicians isolate and cure diseases.[29] Hence, in the context of the antipsychiatry debates, the "medical model" should be taken to connote *individualistic* conceptions of treatment rather than more narrowly construed "biological" ones.

The *medical thesis*, that the function of psychiatry is to treat diseases, is typically argued for on the basis of recent medical advances in the knowledge of the biological correlates of mental disorders, and in their treatment.[30] Historically and sociologically, this has been an overwhelmingly powerful argument: the neurobiological and

---

[28] The most classic exposition of the tenets of the "medical model" of psychiatry is Feighner *et al*. (1972) and Klerman (1978). However, it should also be kept in mind that there is no unambiguous definition or characterization of what the "medical model" is supposed to be; see, e.g., Macklin (1973).

[29] Ironically, this assertion runs counter to Freud's own explicit rejection of the medical status of psychoanalysis. In the postscript to his 1927 book, *The Question of Lay Analysis*, he writes, "I have assumed…that psychoanalysis is not a specialized branch of medicine. I cannot see how it is possible to dispute this…The possibility of its application to medical purposes must not lead us astray" (quoted in Siegler and Osmond [1974, 49]).

[30] The introduction of Chlorpromazine (marketed in the US as Thorazine) into the psychiatric asylum in 1954 for the treatment of schizophrenia is often held as the first major step toward the twentieth century "biological revolution" in psychiatry (Swazey [1974]).

pharmacological advances made by biological psychiatry have, more or less, brought closure to the heyday of antipsychiatric literature. Nonetheless, the argument that medical advances in the understanding and treatment of mental disorders substantiates the medical thesis is straightforwardly question-begging. The discovery that a certain type of behavioral or psychological condition has a biological component does not show that it stems from a *disease* or a biological *dysfunction*; it merely suggests that it has a biological *cause*. For the physicalist, all mental states have physical (and presumably biological) causes. Strictly speaking, all that scientific advances in the understanding of the biological foundations of psychology and behavior show is that psychiatry is becoming progressively more effective in controlling certain types of behavior that are deemed undesirable to have.

A second claim that is often supposed to substantiate the medical thesis is that newer diagnostic and classification systems for mental disorders are more reliable and valid than older diagnostic systems: they are more *reliable* in that the diagnostic descriptions they offer have a higher level of inter-rater agreement, and they are more *valid* in that the diagnoses has greater predictive accuracy.[31] But the question of whether or not current classification systems are more reliable or valid than older ones is orthogonal to the question of what makes the conditions so classified describable as "disorders" or "diseases". The classifications of economic status, ethnicity, and sexual orientation are fairly reliable and valid, although presumably these are not "disease classifications".[32]

---

[31] E.g., Andreasen (1997, 1586).

[32] Presumably, the reason that the improved reliability of current classification systems is often mistakenly used as an argument against antipsychiatry is that some antipsychiatrists, such as Rosenhan (1973), made the unreliability of then-current classification systems, and especially of the concept of "schizophrenia" as used in the United States, the target of their criticism of the profession (see Section 2.1.1 for an elaboration of Rosenhan's criticism). But, conceptually, the relative vagueness or precision of a category has nothing to do with whether it is value-laden or with the nature and source of those values.

The conceptual claim that mental disorders are non-relational properties of individuals is intimately bound up with the medical thesis that the function of psychiatry is to treat diseases. One plausible way of deriving this claim is the following: if a disease is (or stems from) an internal dysfunction on the part of the individual, and an internal dysfunction on the part of the individual a non-relational property of that individual, the view that mental disorders are diseases implies that mental disorders are (or stem from) non-relational properties of individuals.

Despite their intuitive merit, these first two premises are not trivial. The first premise involves a semantic claim that the concept of a disease refers to an internal dysfunction on the part of the individual. However, a significant and long-standing debate within the philosophy of medicine concerns the question of whether the concept of disease refers exclusively to a type of internal state that can, in principle, exist in the absence of any disposition to produce suffering, whether "disease" necessarily refers to a state that disposes its bearer to suffering, or, at the extreme, whether "disease" simply refers to any internal condition that disposes its bearer to suffering, without additional qualification.[33] If the term "disease" merely refers to any internal condition that produces suffering, with no additional qualification, then it is not clear why diseases should involve internal dysfunctions, since intuitively, suffering can occur in the absence of an internal dysfunction. Nonetheless, the conceptual content of the word "disease" will not be pursued any further in this dissertation. It will be stipulated that having an internal "dysfunction" is at least a *necessary* condition for having a "disease", whether or not a

[33] Sedgwick (1981) claims that diseases are just undesirable biological conditions. Canguilhem (1991, 208-9) appears to defend the weaker claim that being a biological condition that disposes its bearer to suffering is a necessary, but not sufficient, condition for being a disease; that is, something would not be a "disease" unless at some historical point someone experienced the condition as an obstacle to his or her goals.

disease is also accompanied by a subjective experience of suffering on the part of the individual who has the condition.[34]

The second premise, that an internal dysfunction on the part of the individual is a non-relational property of the individual, is also a conceptual claim that depends on how the notion of internal "function" and "dysfunction" are explicated. If the notion of an internal "function" is explicated in terms of the contribution of a trait to the relative fitness of the organism that possesses it (i.e., as held by the "propensity theory" of biological functions), then whether a given trait has a biological function depends partly on the organism's environment and hence is not a non-relational property of the individual. On the other hand, if the notion of function is explicated exclusively in terms of the *history* of the functional trait, without making any claims about its *current* contribution to fitness (i.e., as held by the "etiological theory" of biological functions), then whether or not a trait has a function is a property of an individual that is not relative to its current environment but one that supervenes entirely on its structure and history.[35]

1.2.3 Clarification of Thesis in Relation to Antipsychiatry and the Medical Model

Having provided a schematic outline of the classic debate between advocates of the antipsychiatric position and advocates of the medical model, the thesis that will be advocated here can be placed in relation to them. The thesis of this dissertation is that

---

[34] That "causing suffering" is not a necessary condition for having a disease is often argued for on the following basis: intuitively, a person can have a disease without being aware of it and without experiencing physical discomfort (Kendell [1975a, 10]); apparently non-sentient beings, such as plants, can have "diseases" (Boorse [1975, 53]), etc. However, there is no reason that these arguments could not be resolved by restricting "disease" language to sentient beings, rendering "causing suffering" as "being known to have a disposition to cause suffering", and so on. (Perhaps one would never have spoken of "diseases of plants" unless such conditions were disposed to produce suffering on the part of humans; if so, then the fact that plants can have "diseases" is not a good counterexample to the subjectivist view of disease.)

[35] The concept of a biological function will be described in Section 2.3 and Chapter 3; in Chapter 3, a version of the etiological theory of functions will be defended as the only theory capable of satisfying some minimal adequacy criteria that the biological perspective in psychiatry imposes on the explication of function statements.

there is little warrant for the claim that psychiatric disorders stem from biological dysfunctions. This suggests that the norms involved in psychiatric research and classification may be psychological, social, ethical, epistemic, legal, etc., in nature. Hence, this thesis is *compatible* with the weaker sociological thesis that psychiatry is an institution one of whose functions is to explain, predict, and correct certain forms of deviation from widely-held norms of appropriateness that are social in their origin and justification. However, it will also accept the conceptual claims proposed by advocates of the medical model concerning the concept of disease: namely, that diseases *do* entail the presence of internal dysfunctions on the part of individuals, and that such dysfunctions *are* non-relational properties of individuals (that is, they are not relative to that individual's current environment). It will use these conceptual claims to argue *against* the proposition that mental disorders are best conceptualized as diseases.

However, this is not to say that biologically-oriented medical procedures are not appropriate to the amelioration of these conditions. Consequently, the thesis departs from the normative implications that were often drawn, by antipsychiatrists as well as proponents of the medical model of psychiatry, concerning appropriate treatment measures. What should not be neglected, as mentioned at the beginning of the section, is that the debates about the concept of "mental disorder" were often fueled by prior commitments concerning the most appropriate means of ameliorating human suffering – for example, biological, psychodynamic, and community-based treatment methods. Often, these conflicting commitments gave rise to heated conceptual debates concerning the nature of "mental disorder", and these debates were supposed to provide an *a priori* basis for generating acceptance for the favored method of therapy. Szasz (1961), as well as Sarbin (1969), for example, draw heavily upon Ryle's (1949) analysis of "mind" to argue that most forms of psychiatry, whether biologically or psychodynamically oriented,

rest upon a problematic analysis of "mind" as a substantive entity that exists "in the head" and that can provide a substrate for "abnormalities" or "diseases". They then argue that this concept is metaphorical, that "mind" refers instead to a set of practices that are embodied in a matrix of social transactions, and therefore that the purpose of therapy should be to help people to become aware of how their behavior reflects the adoption of social roles that may not be conducive to their well-being. But this is to conflate conceptual analysis and treatment methodology. It would be equally invalid to argue from the physicalist premise that mental states are physical states of the brain to the conclusion that neurological intervention alone is warranted for the alleviation of mental disorders. Certainly, no conceptual analysis of "mind" or "mental disorder" should foreclose the question of effective therapeutic procedures for alleviating suffering or preventing harmful conduct.[36]

1.3 Relation of Thesis to Contemporary Debates: Descriptivism and Prescriptivism

Much of the recent philosophical literature on categories of mental disorder reflects a shift away from the debate between the proponents of antipsychiatry versus those of the medical model; they are more explicitly oriented towards conceptual analysis of medical and psychiatric terms. On the broadest level, the debates concern the question of the relation between facts and values in psychiatric classification and research; more specifically, they concern the extent to which the concept of mental disorder has purely *descriptive* meaning, *prescriptive* meaning, or some mixture of both. Unlike the classical debates, however, there is no implicit assumption that the "scientific status" of psychiatry

[36] Ironically, Matthews (2003) takes up a similar Rylean position to argue for methodological pluralism in psychiatry.

hangs in the balance of the debates, or that, if "mental disorder" contains any prescriptive meaning, then its application would be arbitrary, subjective, or ideologically driven.[37]

The reason that the debate is important is because, if one were to understand how values – here narrowly construed in terms of "prescriptive meaning" – enter into medical and psychiatric concepts, then one would have a better appreciation for the appropriate place of moral discourse in medical theory and practice. For example, if, according to the descriptivist view of "disease" (e.g., Boorse [1977]), whether or not something is a disease does not depend on whether or not its consequences are negatively valued, then moral discourse may occupy a more marginal role within medical theory than it would otherwise, and be limited to standard topics treated in introductory medical ethics courses – the right to die, the nature of competence, the standards for informed consent, and so on. If, on the other hand, whether or not something is a "disease" is partly determined by whether its consequences are negatively valued, then its ascription would presuppose some conception of personal or individual well-being in relation to which disease is an obstruction or hindrance. In this case, at least in principle, the role of moral discourse would be more central to medical decision-making. For example, if there were disagreement concerning whether or not obesity is a disease, or racism is a mental disorder, then it would be helpful to know which types of disagreement could be resolved on fairly straightforward empirical grounds, and which types stem from divergent value-commitments. (This is not to suggest that the latter types of divergence cannot be rationally resolved, or that no other sources of disagreement exist.) The use of conceptual analysis to clarify the role of values in medical theory, then, has substantial practical importance.

---

[37] Unless, of course, one holds a "pure prescriptivist" view as described in the previous section, according to which "mental disorder" possesses no cognitive content and serves, in psychiatric and social discourse, as an epithet for any psychological condition that the speaker disapproves of and wishes his or her audience to disapprove of as well.

The terminology appealed to in these debates stems from R.M. Hare's (1952, 1963) analysis of moral language; hence Hare's usage will be briefly recapitulated.[38] Hare defined the terms "descriptive meaning" and "prescriptive meaning" largely ostensively. The term "red" is paradigmatic of a term with purely descriptive meaning, in that it picks out an empirically discernable property of an object and does so in a way that does not necessarily commit the speaker to any particular value-judgement concerning whether an object's being red is commendable or not.[39] The term "good", however, contains prescriptive meaning in that it essentially performs the function of commendation: to say that something is good is a way of commending that thing, or things that are similar to that thing in some determinate respect. Another way of expressing the idea that a term "essentially performs the function of commendation" is that *any intelligible usage of the term presupposes an act of commendation on the part of the person who utters it*. Thus, a person who says sincerely that "*X* is good", but does not commend *X*, cannot be said to "exhibit mastery" of the English language. Unlike the term "good", the term "courageous" exemplifies both types of meaning, in that it both describes an empirically discernable pattern of behavior or psychological disposition, and has the function of expressing commendation of that property (compare, e.g., "courageous" with "foolhardy" or "reckless", which often agree extensionally but are used to express different value-judgements).

Although the concepts of "descriptive meaning" and "prescriptive meaning" are fairly clear, Hare's definitions of "descriptive term", "prescriptive term", and "evaluative

---

[38] The primary exposition can be found in Hare (1952), Chapter 7; and (1963), Chapter 2. Since the usage is not consistent across texts, the 1963 exposition will be used here. Additionally, reference will be by chapter and section rather than by pagination.

[39] Of course, *any* term can occasionally be used to express commendation; a person who is known to like the color red may exclaim, "Why, it's *red*!" to convey that being red is a praiseworthy quality and to suggest a course of action (Sadler [2004, 29]). However, this does not make "red" an evaluative term in the narrow sense (Ibid., 30) insofar as this commendation is not part of the meaning of the word itself.

term" do not correspond in an obvious manner to the former. This situation has created a significant amount of discrepancy with respect to the way these categories are applied in contemporary literature. In short, according to Hare, a term is a "descriptive term" if it contains exclusively descriptive meaning; a term is a "prescriptive term" if it contains *any* prescriptive meaning (whether or not it also contains descriptive meaning); and a term is an "evaluative term" if it contains both descriptive and prescriptive meaning (1963, §2.8). Consequently, all evaluative terms are necessarily prescriptive terms, but it is possible that there are prescriptive terms that are not evaluative terms.

However – and this is an important caveat – Hare does not believe that moral terms, such as "good", "right", and "ought", contain exclusively prescriptive meaning. Hence they are evaluative terms rather than prescriptive terms. His argument for this is that the purpose of commendation is to establish something as an example for guiding selection (1952, §8.1). For example, to say "Rembrandt is a good painter" is to posit Rembrandt's work, or his creative process, as a model to which aspiring painters should strive. Thus, in order for the statement to fulfill this purpose, the speaker must imply that there are replicable features of Rembrandt's paintings or creative process, and that any painter whose work or creative process satisfies these standards is also a "good painter". Consequently, when one states that "*X* is a good *Y*", one expresses the judgement that *X* has some set of descriptively expressible properties P_1, P_2,…, P_n, and one is committed to the universally quantified statement, "Any *Y* that has properties P_1, P_2,…, P_n, is a good *Y*". Another way of putting this point is that the property of being a "good" example of a kind *supervenes* upon a set of descriptively expressible properties such that two objects cannot differ solely with respect to their goodness. Hence the descriptive meaning of moral terms implies that moral judgements are universalizable (1963, §2.5). The reason for Hare's emphasis on the coexistence of prescriptive and descriptive meaning in moral

language is that it constitutes a weak constraint on the rationality of moral discourse, although one that avoids the so-called "naturalistic fallacy" associated with rejecting the prescriptive meaning of moral terms altogether and defining "good" in terms of a set of descriptively expressible properties (1952, §5.4).

However, it should be noted that this is a very weak constraint on moral discourse, insofar as there is very little descriptive content associated with the term "good" in the way that, e.g., there is specific descriptive content associated with the term "courageous"; for in order to be "courageous" one must, at least, have a psychological disposition to take risks, but no such unique property is implied by being "good". In other words, "courage" is, as Williams (1985) puts it, a "thick" ethical concept (Ibid., 129), because it is relatively constrained by specific descriptive content; "good" would be a "thin" ethical concept because it is relatively unconstrained.

The distinction between descriptive and evaluative terms will allow two major positions on the evaluative character of the term "mental disorder" to be rendered more precisely: if the term "mental disorder" is an evaluative term, then what specific descriptive content, if any, constrains its ascription? In other words, supposing the term "mental disorder" to contain prescriptive meaning, then can virtually any psychological characteristic or phenomenon qualify as a "mental disorder"? Or only very specific sorts of psychological conditions?

The above recapitulation of Hare's schema may seem to belabor the point, but it is important because there is often disagreement concerning the conditions for classifying a given term as "evaluative" or "nonevaluative", and these disagreements tend to distort the debates. Of course, one cannot argue that there is a "correct" usage of the terms "descriptive" and "prescriptive"; one can merely stipulate one's meaning and justify that stipulation in terms of historical usage and current usefulness. In particular, there are two

assumptions that are often made in the debates about the evaluative content of psychiatric nomenclature that will not be made here, since these assumptions have the effect of trivializing the notion of an evaluative term. The first assumption that will not be used here is that "evaluative term" designates a pragmatic, rather than a semantic, category. "Pragmatic" factors for linguistic usage refer to the empirical conditions that motivate the utterance of a term in a given context; "semantic" factors refer to the *meaning* of the term itself or the conditions of its intelligible usage. The second assumption that will not be used here is the assumption that a term is *evaluative* if the conditions of its truth are relative to the values held by some person or another, rather than *specifically relative to the values held by the speaker or one whose values the speaker represents*. Examples of how both assumptions enter into the debates will be provided in order to clarify the relatively narrow scope of the term "evaluative" that will be used here.

An example of the first assumption involves the inference from the fact that a given use of a term is motivated by values (pragmatic) to the claim that the term is evaluative (semantic), that is, that some value-commitment is part of its meaning. Agich (2002), for example, notes uncontroversially that, "[D]isease language arises from a response to the everyday experience of illness, namely the individual experience of feeling bad or of not being able to perform some normal action…"(Ibid., 106). What he believes to follow from this observation is that "[D]isease language is essentially evaluative. It is bound up with evaluative concepts of illness" (Ibid., 107). But to say that disease language is motivated by a negatively-valued experience of suffering (pragmatic) does not imply that such an experience is part of the meaning of the term "disease" (semantic). This inference, if valid, would trivialize the concept of an evaluative term. For example, geometry may have arisen out of the collective desire to exchange and accumulate property, but this does not imply that terms such as "point" and "polygon"

are evaluative. (Geometry may be evaluative in another sense, namely, that geometrical theorems are epistemically normative – in the sense that one *ought* to believe in them – but that is not the sort of value that is typically thought to be involved in the debates concerning the evaluative status of psychiatric terms.)[40]

The second way of trivializing the concept of an evaluative term is to claim that a term is evaluative if it refers to the values held by some person, rather than that it presupposes an act of commendation specifically on the part of the speaker (or one whose values the speaker represents) as a condition of its intelligible usage. Fulford (2000), for example, claims that "rape" is an evaluative term. The reason for this is that absence of voluntary consent is one of the truth-conditions for the correct application of the term "rape" (Ibid., 85), and thus it imputes to the victim a negatively-valued experience. Thus there is a reference to the values held by the victim. However, this fact does not suffice to infer that the term is evaluative in the sense that will be used here, since it does not necessarily impute to the speaker any particular attitude toward rape, or, more generally, to one person's using another as a means to which the latter does not consent. Despite the fact that anyone who could have a neutral attitude about rape exhibits a deficiency in ethical judgement, it is not a contradiction in terms to imagine such a person.

Fulford claims explicitly that the concept of an evaluative term carries no such restriction to the *speaker's* values. Yet, if there is no such implication, then virtually *all* psychological description is evaluative, in that any factual description of a person's emotions, desires, values, or preferences is "evaluative". Psychiatric classification would

[40] Sadler (2004) draws a helpful distinction between the fairly encompassing idea of value-ladenness and the relatively narrow idea of a "value-term". The notion of "value-ladenness" can be applied to virtually any linguistic or conceptual entity – words, sentences, discourses, and theories – that is associated with values, where a value is a concept that is used to impute praise or blame to something and to guide action and is indifferent between semantic and pragmatic considerations. The notion of a "value term" is much more narrowly construed to refer to individual words or expressions that have prescriptive meaning. In this sense, one may say of, e.g., conservation biology that it embodies "value commitments", and that it is a "value-laden" science even if no particular term in its specialized nomenclature is a "value term".

be "evaluative" merely because it deploys psychological descriptions. This would seem to trivialize the debates concerning whether psychiatric nomenclature is evaluative. By contrast, in the sense of the term to be used in this context, to describe a person as a "victim" of rape *is* evaluative, since the use of the term "victim" necessarily imputes to the speaker the belief that the person was wrongly violated, and that freedom from such transgression is inherently good. (Of course, it may be that the notion of "voluntary consent" *is* evaluative, if by stating this, the speaker imputes to the person the "competence" to consent, and if the notion of "competence" is only definable in terms of epistemic, legal, or ethical norms. But that was not the point that Fulford was making). Hence, to say that "mental disorder" is evaluative is to say that the intelligible usage of the statement "*X* is a mental disorder" presupposes an act of commendation on the part of the speaker (or the community that the speaker purports to represent) – for example, that *X* is prima facie undesirable to have, or that it produces negative consequences, or that it ought to be treated, etc.

Having provided an overview of the conceptual apparatus that informs the debate, as well as some common assumptions that will not be made here, some of the major contrasting positions on the evaluative content of the concept of a mental disorder can be described. Interestingly, there are very few *pure* prescriptivists or *pure* descriptivists with respect to the meaning of "mental disorder" or other psychiatric or medical terms. Most of the participants assume that the term *is* evaluative, but disagree about the precise nature of the descriptive constraints that are imposed upon it. Hence the debates cannot accurately be glossed in terms of the opposition between "descriptivists" and "prescriptivists", or between "naturalists" and "normativists", as is often done.

One view is that "mental disorder" is a relatively constrained by specific descriptive content. According to one prominent analysis, given by Wakefield (e.g.,

1992a; 1992b; 1999a), "disorder" *as such* can be analyzed as "harmful dysfunction". For Wakefield, the notion of "harm" contains a prescriptive component. Presumably, for a person to judge something to be "harmful" is for that person to judge that it has the capacity to undermine or destroy something that he or she deems worthy of preservation, something that ought to be protected, and so on. Insofar as there are few, if any, intelligibility constraints upon what sorts of items a person may deem worthy of preservation, then the judgement that something is harmful is relatively unconstrained by any specific descriptive content. However, according to Wakefield, the concept of "dysfunction" contains *only* descriptive meaning. It refers to the inability of a biological trait to perform its function, where a "function" of a trait can be defined, in turn, exclusively in terms of the activity that it was selected for by natural selection to perform. Hence according to Wakefield's view, the term "mental disorder" is an evaluative term that is relatively constrained by descriptive content. If this is correct, then even though one should acknowledge the relevance of value judgements to psychiatric research and classification, one should not be a radical antipsychiatrist and assume that mental disorder classification is merely an expression of socially-conditioned value judgements. This position will be developed in more detail in Chapter 2; it is an elaboration of Klein's (1978) analysis of "illness" as an involuntary impairment of an evolved function that warrants the "sick role" (Section 2.3.3).

Wakefield's position, however, prompts a closer analysis of the allegedly descriptive status of "dysfunction" (and the corresponding concept of function). Although Wakefield is a descriptivist with respect to the term "dysfunction", others have proposed that the concept of dysfunction itself contains prescriptive meaning. Moreover, philosophers who hold that the term "dysfunction" is evaluative can be divided, like those who believe the term "mental disorder" to be evaluative, into those that hold

"dysfunction" to be relatively constrained or unconstrained by specific descriptive content.

Sedgwick (1981), for example, holds that the concept of "disease" is relatively unconstrained by specific descriptive content; his analysis, however, can easily be extended to the notion of "dysfunction". For Sedgwick, to say that something is a disease is to say merely that it is an undesirable biological condition: "Outside the significances that man voluntarily attaches to certain conditions, *there are no illnesses or diseases in nature*" (Ibid., 121). Presumably, if one were to extend his analysis to the concept of dysfunction, Sedgwick would hold that the term "biological dysfunction" merely refers to a biological condition that the speaker deems undesirable to have. Consequently, to append the notion of "dysfunction" to the notion of a "biological" condition does not amplify the descriptive content of the latter. Hence it would be relatively unconstrained by specific descriptive content. This view, of course, would obviously undermine the point of invoking the notion of a biological dysfunction to justify the ascription of a mental disorder: to say that a disorder is a "harmful dysfunction" would just be to say that it is harmful and that it represents a condition that is undesirable to have. Similarly, Moore (1978, 103) seems to hold that a person's conception of "functional organization" is structured around his or her conception of well-being and thus that there is no value-neutral fact about what constitutes the proper functioning of a trait (also see Toulman [1975, 60-62]; Engelhardt [1976, 133-4], and Erde [1979, 44] who hold similar views).

Others hold that "function" and "dysfunction" are evaluative terms that are relatively constrained by specific descriptive content. An example of a person who holds this position on the concept of function is McLaughlin (2001) (although Bedau [1991, 1993] may be interpreted as holding a similar position; also see Megone [2000]).[41] Like

[41] In the philosophy of psychiatry, one of the most prominent advocates of the view that "function" is evaluative is Fulford (1989, 1999, 2000). However, Fulford's view will not be elaborated here because he

Wakefield, McLaughlin believes that in order for a biological trait to have a function, it must have historically contributed to its own persistence or reproduction. Thus, "function" has some descriptive meaning. However this is not sufficient: its activity must also be judged to "benefit" the organism in question (Ibid., 191). To judge that an organism "benefits" from the activity of a trait is to judge that the organism, unlike a simple autocatalytic reaction that perpetuates itself across time, is the sort of thing that can have a "good". Hence to say that the function of the heart is to beat is to say the heart's beating *benefits* the individual that possesses it *and* that it has contributed to the reproduction of the organism precisely because it has so benefited the organism. In other words, according to this view, "function" contains prescriptive as well as descriptive meaning, because to say that performance of the function is beneficial to the organism is to imply that it serves as a means to an end that the speaker considers good, i.e., survival or reproduction.

A critical point that is often ignored in the debates concerning whether "function" is an evaluative term is that there are several different concepts of "biological function" in circulation in biology. At times, "function" simply means something along the lines of "characteristic activity of a structure" (e.g., Bock and von Wahlert [1965]). In this sense, both "circulating blood" as well as "producing oscillations in air pressure that can be detected through a stethoscope" can be referred to as functions of the heart. It is plausible this liberal construal of "function" is purely descriptive. However, it is also plausible that this notion of function is not strong enough to explicate the sense in which pumping blood is the "proper function" of the heart and making beating sounds is not (e.g., that the latter represents an "accidental by-product" of the heart's proper functioning). Similarly,

---

does not offer any particular analysis of the notion of "function"; rather, he argues that close attention to the analogy between artifact "functions" and biological "functions" reveals that *any* particular explication of "function" that introduces normativity into the biological realm must have prescriptive meaning.

it does not seem to provide any explication of the locution that the heart can "malfunction" when it does not pump blood. As described at the beginning of this chapter, one prominent notion of "function" is normative: the statement "the function of *X* is *Y*" does not describe the current activity or disposition of *X*; rather it says of *X* that it is "*supposed to*" do *Y*, or that the *purpose* of *X* is to do *Y*, and that it is *malfunctioning* (e.g., functioning poorly) otherwise. (Often the "litmus test" for whether or not a concept of function is normative is whether it supports the inference that a part can "malfunction".) Hence, the debate about the evaluative content of the term "biological function" can be made more precise by saying that, according to the evaluativist position, any notion of function that is *normative* must possess *prescriptive meaning* (in that the speaker must commend the functional activity, or believe that the outcome of such activity is good – e.g., survival), and that this prescriptive meaning explains its normativity. The descriptivist position is that the normative content of function statements does not entail that the term contains prescriptive meaning; that is, the truth of the judgement that a part is "functioning properly" or "malfunctioning" contains no implicit reference to the values of the person who so judges.

Having provided an overview of the current framework for the debates on the value-ladenness of psychiatric classification, the thesis of this dissertation and the argumentative strategy it employs can be placed in relation to them. Although it does not offer or defend any particular analysis of the concept of mental disorder, it will adopt, as a working definition, an explication that is consistent with the biological perspective in psychiatry; namely, that "mental disorder" is an evaluative term that refers to a *failure of social-role functioning* caused by an *internal dysfunction* on the part of the individual (see Section 2.2). Insofar as the notion of *failure of social-role functioning* contains prescriptive meaning, then the definition adopted here implies that the term "mental

disorder" is evaluative. However, the dissertation *will* provide and defend an explication of "biological function". According to the explication that will be proposed here, the concept of function is a descriptive one, and it is also normative. It is descriptive in that to say that, "the function of *X* is *Y*", does not necessarily impute to the speaker any particular attitude about *X*'s doing *Y* or its outcome. It is normative in that to say that, "the function of *X* is *Y*", is to say that *X* is *supposed* to do *Y* and that *X* is capable of *malfunctioning*. Accordingly, the definition of "mental disorder" that is adopted here implies that it is an evaluative term that is relatively constrained by descriptive content.

To the extent that this dissertation defends an explication of the concept of function according to which the concept of function (dysfunction) is descriptive, it is in basic agreement with Wakefield's view. However, the use to which the concept of function will be put departs significantly from the analytically-oriented framework of the debates, for it will concern two questions that have scarcely been asked in the literature:

> (i) given an explication of the notion of a biological function, how can biological functions, or dysfunctions, be empirically identified? Providing a definition is not sufficient to resolve any substantive issues if the definition does not lend itself to the construction of appropriate empirical indices that warrant its application; and
>
> (ii) given the definition of "biological function" as well as the empirical indices for its warranted application, is there any reason to believe that those conditions that are commonly identified as mental disorders – and in particular, those that are standardly targeted by biologically-oriented researchers – in fact stem from biological dysfunctions on the part of the individual? Furthermore, is there any reason to believe that they soon will be shown to stem from biological

dysfunctions? The answer that the dissertation gives to both of these questions is that there is no reason to believe that these conditions stem from biological dysfunctions, or that they will soon be shown to do so.

As a consequence, although the dissertation attempts to avoid some of the conceptual issues concerning the precise explication of "mental disorder", it does have an important bearing on the debates. First, if "mental disorder" is explicated in such a way that it necessarily refers to an internal dysfunction on the part of the individual, then there are few, if any, mental disorders. Hence, supposing that there *are* mental disorders, such an analysis does not constitute a good explication of the term. To the extent that pursuing the correct analysis of "mental disorder" is a worthwhile task, then those who pursue this task ought to seek out a different basis for explicating the concept – for example, via a psychological, phenomenological, moral, social, or legal analysis. In short, then, the conclusion that the dissertation draws is in basic agreement with the view of, e.g., Fulford (1994) and Sadler (1999), both of whom argue that the notion of "dysfunction" is of limited value in explicating the normative dimensions of mental disorder ascriptions and of limited relevance to the practical demands of mental disorder classification and diagnosis.

1.4 ARGUMENT OF THE DISSERTATION

The next three paragraphs provide a short overview of the argument of the dissertation. The remainder of the section will provide a more detailed description of that argument.

The argument will begin by assuming that there is a notion of "biological function" that is appropriate to the context of psychiatry, and that once this concept of biological function is understood, then the "biological dysfunction" position – that the

origin and justification of the standards or norms of appropriateness that are appealed to in psychiatric categories are explicable purely on the basis of biological considerations – will be given a precise formulation. On the basis of this formulation, a clarification of the nature of the *evidence* that would have to be marshaled in order to warrant its application in any given case can be provided. Thus the concept of biological function that will be proposed will allow the question of the origin and justification of the norms and standards of appropriateness appealed to in psychiatric research to be rendered amenable to empirical research, rather than exclusively to conceptual analysis.

This notion of biological function will then be used as a framework primarily for interpreting experimental work on the biological basis of schizophrenia. It will show that the current evidence and hypotheses concerning the biological basis of schizophrenia fail to show that it results from a biological dysfunction on the part of the individual, though it may have one or several biological causes. The reason that the dissertation specifically focuses on schizophrenia is that the heterogeneous set of conditions that fall under that category are paradigms of lay or colloquial usage of "mental disorder" itself, or its more pejorative associations: "madness", "insanity", and "craziness" (e.g., see Smith [1982, 13]; Heinrichs [2001, 3]). This implies that the norms of appropriateness appealed to in what is currently considered to be a central target of biologically-oriented research in psychiatry, as well as a paradigm case of "mental disorder", do not currently have biological justification.

But the sorts of considerations relevant to establishing that schizophrenia does not necessarily stem from a biological dysfunction are equally applicable to biological research associated with other mental disorders, such as bipolar disorder or antisocial personality disorder. The problem is that psychiatrists too often assume that just because a mental disorder is associated with some biological abnormality, then it must result from

a biological dysfunction. Since a careful evaluation of some of the biological abnormalities associated with schizophrenia shows that this is not necessarily the case, it raises a legitimate question: why should it be the case for the other major mental disorders? Although the dissertation leaves open the possibility that schizophrenia – and other mental disorders – *may* stem from biological dysfunctions, it also shows that there is little warrant for this claim.

The remainder of this section will provide a more detailed account of the argument. The argument will begin by analyzing the concept of biological function. It will be argued that there are two conditions of adequacy on any explication of the concept of biological function that is appropriate to the psychiatric context. The first is that it should lend itself to the explication of a corresponding notion of "dysfunction". To say of a biological entity, however, that it is "dysfunctional" is to apply a standard or norm for evaluating its activity. Thus, in accordance with the "biological dysfunction" position, the origin and justification of this standard must be explicated without presupposing unanalyzed norms concerning the "inappropriateness" of the psychological or behavioral conditions that the biological entity produces. The second is that whether or not an item, *X*, is dysfunctional should not be determined on purely externalist grounds – that is, exclusively by changes in the environment that have no effect upon the characteristic structure or activity of *X*.

It will then be argued that any concept of biological function that is that is appropriate for the psychiatric context – that is, any notion of function that can fulfill both adequacy conditions – must be an etiological one. More specifically, the conditions for establishing the claim that "biological item *X* is dysfunctional because it is not performing activity *A*" must make some reference to *X*'s (or objects of the same type as *X*) *selection history*, that is, its actual past history of performing *A*, where this past history

of performing *A* is partly responsible for the fact that *X* has been selected for and thereby has been maintained in a population to the present day. A simple example is the following: one of the truth conditions for the claim that "the heart is dysfunctional because it is no longer circulating blood", is the fact that one of the reasons that hearts have persisted to the present day is that they circulated blood in the past, that this activity partly explains why hearts were selected for by natural selection, which, in turn, partly explains why most organisms today have hearts.

However, it will be pointed out that etiological theories of function, as they are conventionally articulated, suffer from a significant empirical problem, which is that many explications of the concept of biological function assume, either as a matter of definition or of empirical fact, that natural selection operating over an evolutionary time frame is the only type of "selection process" that there is. The problem that this raises is that we often have little or no knowledge about the evolutionary history of the psychological mechanisms investigated by psychiatry – despite the speculative claims of "evolutionary psychiatry" or "evolutionary psychology" – or even of the evolutionary history of their neurobiological correlates.[42] This suggests that, in the context of psychiatry, the locution that a given part of the brain is "functioning properly", or "malfunctioning", is not empirically warrantable on the basis of the methods currently available to the natural scientist, and must therefore rely on norms of appropriateness that are not biologically justified.

The empirical problem of function ascriptions will be resolved by generalizing the notion of a "selection process" embodied in the etiological theory of functions beyond natural selection operating at the level of the individual. For example, immunological

[42] Thus the dissertation will contain an implicit critique of current "evolutionary psychology" (see e.g., Buss [1999]) as well as what refers to itself as "evolutionary psychiatry" (see, e.g., Nesse [1999]; Stevens and Price [2000]).

selection processes, as well as synaptic selection processes, are two selection processes that operate below the level of the individual, and their operation is, in principle, empirically discernable by the methods currently available to the natural scientist. According to synaptic selectionism, the formation of synaptic structures involves a (potentially iterative) competitive process in which, after an initially (partly) random proliferation of new synapses, those that are activated are retained and those that are not are eliminated. The connection between synaptic selection and natural selection was originally proposed by Wilhelm Roux in 1881 (Jacobson [1991, 231]) and has been advanced more recently by Jerne (1967), Changeux and Danchin (1976), and Edelman (1987), as an alternative to "chemotaxic" theories of synaptic formation, according to which pre-existing chemical markers guide the process of axonal growth toward specific terminals in the absence of selective activity, and "constructivist" theories, according to which synaptic growth is stimulated in a non-selectionist manner by neural activity. To the extent that synaptic structure formation can be adequately modeled as a selection process then it should be recognized as a function-bestowing process. Moreover, to the extent that synaptic selection processes are implicated in mediating specifically "psychological" activities such as learning, then this provides a biological reason to consider experience as a process capable of bestowing new functions onto biological entities. In this way, the array of empirical material that is relevant to establishing claims about the biological function of an entity will be extended to include evidence that is in principle empirically discernable given existing methods employed in the natural sciences.

Once equipped with a theory of biological function that can specify precisely the conditions under which an entity can be dysfunctional, the theory will be applied to current research on the biological basis of schizophrenia. It will look at two approaches to

schizophrenia in particular – a neurochemical approach and a neurodevelopmental approach. It will show that the nature of the neurochemical or neurodevelopmental abnormalities that may be associated with schizophrenia do not necessarily imply that it stems from a biological dysfunction. It could be that the specific neurobiological abnormalities associated with schizophrenia represent a brain that is unable to perform its proper function because of abnormal environmental circumstances, or, a brain that is functioning normally. Because of the fact that the sorts of considerations that were relevant for establishing this conclusion are applicable more generally to biological research on mental disorders, then the conclusion should be generalizable. The problem, as stated above, is that many psychiatrists simply assume that just because there are biological abnormalities associated with mental disorders, that means that they are caused by biological *dysfunctions*. Once this assumption is removed, then there is little warrant for claiming that mental disorders are, in fact, caused by biological dysfunctions – although the possibility remains that they may be.

1.5 CHAPTER OVERVIEW

After the introduction (Chapter 1), the purpose of the Chapter 2 is to motivate the analysis of the concept of biological function that will occupy Chapters 3 and 4. It will describe the historical context that prompted psychiatrists in the 1970s and 1980s to critically evaluate the concept of mental disorder and to analyze it in terms of an "internal dysfunction". On this basis it will impose two criteria of adequacy upon the explication of any such notion in the psychiatric context. It will also argue that attempts to analyze "mental disorder" along these lines by psychiatrists in the 1970s largely failed to identify precisely the nature of the evidence that warrants the ascription of a "dysfunction" to an inner state of a person diagnosed as having a mental disorder. The reason for this failure is that the concepts of "function" and "dysfunction" were not adequately defined.

Chapter 3 turns to a broad spectrum of current philosophical analyses of "biological function" and provides an introduction and overview to that literature. It will argue that etiological theories of function are both necessary and sufficient for satisfying the two adequacy criteria.

Chapter 4 will endorse a specific version of the etiological theory as particularly appropriate for the psychiatric context. It will argue that the appropriate etiological theory should refer to the operation of natural selection, but it will point out that selection processes do not only operate over an evolutionary time scale. They also operate over the lifetime of the individual. For example, synaptic structures in the brain undergo a type of "natural selection", and because of that, they can have functions, even if they do not undergo natural selection in the evolutionary sense. This allows one to provide detailed, empirical criteria for deciding whether or not a given synaptic structure has a biological function, or whether it is dysfunctional.

Chapter 5 will argue that current biological evidence concerning the etiology of schizophrenia (according to neurochemical and neurodevelopmental models) fails to show that it results from a biological dysfunction on the part of the brain (though it may have a biological cause). It will conclude the dissertation by suggesting that psychiatric disorders in general should not be conceptualized as stemming from biological dysfunctions on the part of the brain unless there is specific evidence for this conclusion other than the existence of biological abnormalities. Chapter 6 will provide a brief discussion of the bearing that this conclusion has for how mental disorders are conceptualized, both for the psychiatric practitioner as well as the layperson.

Chapter 2: From Mental Disorders to Internal Dysfunctions

The purpose of this chapter is to motivate the explication of the notion of an internal "function" and "dysfunction" – that is, to show why it is important for these terms to be defined in the psychiatric context. The first section will provide the historical context for the interest taken by American psychiatrists in the 1970s in formally defining the concept of mental disorder (Section 2.1). In the 1970s, the American Psychiatric Association (APA) was beset by two crises concerning its identity as a scientific and therapeutic discipline. Both of these crises were clearly exhibited in a series of debates that took place within the APA on the psychiatric status of homosexuality (Section 2.1.1). As a response to these crises, many psychiatrists came to believe that a definition of "mental disorder" would provide a principled set of criteria on the basis of which such debates could be rationally resolved. More generally, they believed that such a definition would allow mental disorders to be clearly distinguished from socially-disvalued psychological or behavioral conditions. In accordance with the medical orientation of psychiatry, many psychiatrists attempted to define "mental disorder" partly in terms of an "internal dysfunction" on the part of the individual (Section 2.1.2). For example, "mental disorder" could be defined, as it is in DSM-III (APA [1980]) and later editions, as a psychological or behavioral syndrome that is caused by an internal dysfunction and that produces negative consequences for the person who has it. This definition, then, prompts the question of what, precisely, an internal "dysfunction" consists of.

Section 2.2 will provide a philosophical critique of the characterization of "mental disorder" that appears in the DSM-III and later editions (Section 2.2.1). It argues that, in the absence of any explicit definition of "internal dysfunction", one has no way of assessing whether the definition of "mental disorder" offered can successfully fulfill its

purpose of differentiating mental disorders *proper* from any negatively-valued psychological and behavioral condition (Section 2.2.2). This is why it is important to define the term. It then identifies two criteria of adequacy that any definition of "function" should satisfy if it is to perform the role of delimiting mental disorders from other conditions that do not fall under the purview of psychiatry.

Section 2.3 will examine three different attempts made by psychiatrists to define a notion of "dysfunction": the "biological disadvantage" definition (Section 2.3.1); the "operational" definition (Section 2.3.2); and the "evolutionary definition" (Section 2.3.3). It argues that the first two definitions fail to satisfy the adequacy criteria, and that although the third one does, in principle, satisfy those criteria, its formulation does not lend itself to the construction of appropriate empirical indices to warrant its application in any given case. Nonetheless, this shortcoming will be rectified in later chapters. This critique sets the stage for Chapter 3, which provides a more refined classification and analysis of different definitions of "function" and "dysfunction".

2.1 HISTORICAL MOTIVATION FOR DEFINING "MENTAL DISORDER"

This section will describe the historical context of American psychiatry in the 1970s and the crises of professional identity that it faced. It will explain how and why many psychiatrists came to believe that an explicit working definition of "mental disorder" would help to resolve those crises. In particular, these crises led to the belief that "mental disorder" should be defined partly in terms of an *internal dysfunction* on the part of the individual, where the notion of internal dysfunction was intended to signify a non-relational property of a person's psychological, behavioral, or biological condition that is conceptually independent of the undesirable consequences it produces, and that falls within the domain of medicine. "Mental disorder" could then be defined in terms of

an internal dysfunction that, in turn, produces negative consequences for the person who has it.

These crises of professional identity were amplified by the fact that, sociologically, the APA was at a critical juncture with respect to the clinical and therapeutic orientation of the psychiatrists that constituted its membership, as well as its fundamental approach to mental disorder classification and diagnosis. On the one hand, the primarily psychodynamic orientation of the APA constituency was beginning to yield to a more biologically and behaviorally oriented constituency (see below). On the other hand, the then-most-recent edition of the APA's official diagnostic manual, DSM-II (APA [1968]), was undergoing a substantial reconfiguration in form and content that would eventually culminate in the publication of DSM-III (APA [1980]). (See Section 2.1.2, under "The Charge to Define 'Mental Disorder'".) These two transformations are not independent of one another, for the APA task force responsible for preparing the DSM-III was composed primarily of research psychiatrists with a strong biological or behavioral orientation. Consequently, the form and content of the DSM-III reflects this orientation in that it introduces explicit and precise membership criteria for each mental disorder category and, to the extent possible, these criteria are based on observable variables (or at least on those variables that could be easily gleaned from a brief psychiatric consultation). These criteria were introduced to replace the sorts of diagnostic descriptions found in DSM-II, many of which were couched in the jargon of specific psychodynamic etiological hypotheses. Thus the transition from DSM-II to DSM-III reflects the broader shift in the professional orientation of the APA.

2.1.1 Two Crises of American Psychiatry

Psychiatry is often said to have undergone two crises in the 1970s, an external crisis and an internal one. The first is the crisis of legitimation (Moore [1978, 90]; Wilson

[1993, 404]). In the heyday of antipsychiatry, psychiatry was often depicted as having the tacit function of regulating social deviance under the pretense of being a "medical" establishment (see Section 1.2 and references therein on the antipsychiatry tradition). This placed its legitimacy as a scientific and therapeutic enterprise into question. Many of the challenges to this legitimacy stemmed from the widely shared belief that psychiatry possessed no objective and principled basis for distinguishing between disorders and non-disorders. In the absence of such standards, it was charged, social and political power invested in psychiatry could easily lend itself to oppressive ends (Wilson [1993, 404], Dain [1994, 430]). Commercially successful films such as *One Flew Over the Cuckoo's Nest*, replete with harrowing images of electroconvulsive therapy and the effects of lobotomy, served to reinforce this association between psychiatry and violent social repression amongst the public in general (Shorter [1997, 275]). Hence the legitimation crisis was "external" in that it involved the public and academic perception of psychiatry (and of the mental health professions more generally) as having only a dubious entitlement to the social and legal authority that it in fact possessed.

The second crisis concerned the disciplinary position of psychiatry with respect to the other mental health professions, and in particular, with respect to clinical psychology. Specifically, it concerned the question of how psychiatry should distinguish itself from the other mental health professions, both in terms of its object and method.[43] Throughout

[43] Moore (1978, 87) refers to this as a "jurisdictional" concern; also see Blashfield (1984, 65) on the problem of the "professional jurisdiction" of psychiatry. However, phrasing the problem as a "jurisdictional" concern is misleading since it suggests that psychiatrists were exclusively concerned with defining their special *object* of treatment and distinguishing it from those treated by clinical psychologists, e.g., to claim jurisdiction over the "psychotic" rather than "neurotic" disorders. Most psychiatrists would probably agree that what distinguishes psychiatry from clinical psychology is not a different *object* of treatment – since, after all, both are concerned with the treatment of mental disorders. Rather, it consists of a different *method* of treatment – e.g., psychiatrists use those methods that require a medical degree to employ successfully. Hence, in the following, the more unwieldy expression, "crisis of disciplinary position", will be used, rather than "crisis of professional jurisdiction", because the former is broader than the latter insofar as it suggests that the problem concerns where psychiatry "fits" into the system of theoretical and practical scientific disciplines.

the 1970s, the APA was beset by significant internal tension, which can be explained partly in terms of its practitioners' competing conceptions of what sort of discipline psychiatry is. In particular, these competing conceptions often placed psychodynamically-oriented practitioners in opposition to non-psychodynamically-oriented (e.g., behaviorally and biologically oriented) practitioners regarding specific APA proposals. For example, these differences were reflected in a heated dispute concerning the proposal to remove the word "neurosis" from the DSM-III, which was slated for publication in 1980. Those behaviorally and biologically oriented psychiatrists who opted for its removal argued that the use of "neuroses" to identify a diagnostic class involves a commitment to questionable psychodynamic etiological theories and that this commitment runs counter to the atheoretical and descriptive spirit of the DSM-III; whereas those psychoanalysts who defended its retention believed that the validity of those etiological hypotheses had been proven beyond doubt by decades of clinical experience (Bayer and Spitzer [1985]).

This tension was not only internal to the APA. Much more generally, American psychiatry (as represented by the APA) as a whole stood in a certain tension with the field of psychology (as represented by the American Psychological Association) concerning their disciplinary boundaries and areas of relative expertise. By the late 1970s, clinical psychologists were routinely accusing psychiatrists of illegitimately encroaching upon their area of professional expertise by "medicalizing" fairly common sources of psychological malaise that one could adequately resolve without having a medical degree (e.g., Schacht and Nathan [1977]; Garmezy [1978]; McReynolds [1979]). Hence the problem of the disciplinary position of psychiatry constituted a crisis that was internal to the mental health professions as a whole in that it involved an attempt to reach a consensus amongst psychiatrists of different therapeutic orientations, and more

generally, amongst psychiatrists and other mental health practitioners, concerning the nature of the conditions that psychiatry has the special responsibility to treat, and the methods with which it is especially equipped to treat them. The need for psychiatrists to define their area of professional responsibility was particularly pronounced because it coincided with the nearly decade-long process of drafting the DSM-III, which was widely understood to constitute an "official position statement" of the APA concerning the nature of mental disorders.

On the surface, the two crises are related in the following sense: one plausible solution to the legitimation crisis would be to specify explicitly a principled and objective set of criteria for delimiting disordered from non-disordered psychological or behavioral conditions. For all intents and purposes, this would be tantamount to providing a definition of "mental disorder". One could potentially use such a definition to delimit the specific domain of psychiatry and thereby establish the disciplinary position of psychiatry within the mental health professions. In other words, one could define "psychiatry" as that discipline that has the goal of treating the conditions specified by the definition of "mental disorder". This is, in fact, what politically prominent members of the APA attempted to do. The problem with this strategy, of course, is that mental disorders do not constitute the exclusive domain of psychiatry. Consequently, even if a consensus were to be achieved amongst psychiatrists concerning an appropriate working definition of "mental disorder", such a definition would almost inevitably embody the limitations of the way in which psychiatrists tend to conceptualize the object of their field, and for that reason would not find general acceptance amongst mental health professionals. This, as will be seen, is primarily why the attempt to reach consensus on a definition of "mental disorder" did not come to fruition.

The debates that erupted in the early 1970s within the APA on the psychiatric status of homosexuality clearly exemplify both of these crises. They exemplify the legitimation crisis insofar as labeling homosexuality a mental disorder was seen by many people as uncomfortably occupying the borderline between science and ideology. They exemplify the crisis of disciplinary position insofar as they forced psychiatrists of different persuasions to attempt to articulate the concept of mental disorder itself and thereby to articulate their contrasting perspectives on the very nature of psychiatry and its mission. Hence, the next two subsections will discuss how these two crises manifested themselves in the debates on homosexuality and how the debates, in turn, stimulated the attempt to provide a rigorous definition of "mental disorder".

Homosexuality and the Legitimation Crisis

The debate within the APA that lasted throughout the 1970s on the psychiatric status of homosexuality became the primary symbol of the legitimation crisis, since it was easy, although simplistic, to reconstruct the debate as one that opposed psychiatrists who were intolerant of difference in sexual orientation (and thereby sought to extend the concept of mental disorder to include homosexuality), on the one hand, and gay rights activists who sought liberation from the stigmatizing and ideological burden of being thought "diseased", on the other (e.g., see Gold's contribution to Stoller *et al*. [1973]).[44] This ongoing debate, moreover, was widely reported in major newspapers given the publicity that gay rights activists brought to their cause. Starting in 1970, activists staged protests at annual APA conferences and disrupted presentations by psychoanalysts such as Charles Socarides and Irving Bieber who strongly advocated the view that homosexuality represents the outcome of a pathological form of sexual development

[44] Stoller *et al*. collects a large number of position statements by different psychiatrists and other interested parties on the psychiatric status of homosexuality; therefore most of the references will be to that collection.

(Kirk and Kutchins [1992, 82]). Thus, insofar as the troubling DSM-II (APA [1968]) classification of homosexuality as a mental disorder could be held to represent a broad consensus amongst psychiatrists, psychiatry appeared to have positioned itself on the wrong side of a socially progressive cause.

The debates on homosexuality, however, were not the only instance in which the legitimation crisis publicly came to the fore. Several studies undertaken in the late 1960s and early 1970s had the consequence of undermining the public and professional confidence in the profession's ability to identify objectively persons with mental disorders. In particular, they suggested that the label of "schizophrenia" had become so widely deployed in the United States as to become empty of meaning, and this gave rise to the view that such labeling was governed by the arbitrary and subjective whim of its individual practitioners. One of these studies in diagnostic reliability (Katz *et al*. [1969]), for example, analyzed differences in diagnostic practices between American and British psychiatrists. The researchers presented a filmed psychiatric interview to an audience of 42 American and 32 British psychiatrists, in which a young woman evinced anxiety and depression, and complained of frustration relating to the quality of her career and her interpersonal relationships. In response, one third of the American psychiatrists submitted a diagnosis of schizophrenia, although none of the British psychiatrists did the same.

Motivated by such results, a group of researchers initiated the US-UK Diagnostic Project (Cooper *et al*. [1972]; also see Kendell *et al*. [1971] for a brief overview of methods and results) to evaluate systematically such diagnostic differences by comparing hospital admission rates as well as through videotape studies of the sort described above. According to its authors, the results indicate that:

> [T]he concept of schizophrenia held by psychiatrists in the New York area is much broader than that held by London psychiatrists and embraces substantial

> parts of what the latter would regard as depressive illness, neurotic illness or personality disorder, and almost the whole of what would be regarded in London as mania (Ibid., 124).

At least one of the authors concludes from these results that:

> [T]he New York concept of schizophrenia is not a useful one and it likely to inhibit fruitful research if it is widely adopted. We say this largely because the concept has, through the accretion of subsidiary concepts like schizo-affective schizophrenia and pseudoneurotic schizophrenia…become so all-embracing that it is close to becoming a synonym for functional mental illness, a sort of twentieth-century reincarnation of Zeller's *Einheitspsychose* (Ibid., 125).

Interestingly, the most well-known and provocative study that suggested that the American concept of schizophrenia was deployed in a haphazard and arbitrary manner was much simpler than the US-UK research study. It was conducted by David L. Rosenhan, a sociologist, and published in *Science* in 1973 (Rosenhan [1973]). His experiment involved eight normal volunteers who sought psychiatric consultation at different times with twelve different hospitals across the US. They falsely reported a single symptom: being disturbed by a voice that often repeated a certain phrase, such as "dull", "thud", or "empty". They were admitted as patients by eleven of the twelve hospitals with a diagnosis of schizophrenia (the twelfth admitted one with a diagnosis of manic-depressive disorder), and, once admitted, the volunteers no longer made any complaints about their putative condition. The volunteers were retained on in-patient status from seven to 52 days (the average length of stay being 19 days), and those who had been diagnosed with schizophrenia were eventually released with a diagnosis of "schizophrenia in remission".

The purpose of this research was to show, as Rosenhan provocatively stated it, that "we cannot distinguish the sane from the insane in psychiatric hospitals" (Rosenhan

[1973, 257]). Rosenhan believed that if psychiatrists could reliably distinguish the "sane" from the "insane" then the pseudo-patients would not have been diagnosed as having a mental disorder, even if they were to have gained provisional admission.[45] Despite its methodological shortcomings and rhetorical excesses (see Spitzer [1975]; Millon [1975] for criticism of Rosenhan's study; see Rosenhan [1975] for a reply), the prominence of the journal in which the study was published ensured its rapid dissemination amongst the educated public.

Homosexuality and the Crisis of Psychiatry's Disciplinary Position

The debates on homosexuality, however, do not merely function as a prominent symbol of the legitimation crisis. More importantly, the debates stand as an equally salient symbol of the crisis of psychiatry's disciplinary position, because it forced psychiatrists to bring their competing conceptions of what mental disorders are to the surface, and to engage explicitly in the attempt to clarify the boundaries of the pathological – and, by implication, the nature of the conditions that they have the professional responsibility and competence to treat. This can most clearly be seen in the various and often incommensurable criteria that were offered by different psychiatrists at the time to justify the inclusion, or exclusion, of homosexuality as a distinct diagnostic category.

The difficulty of contriving a principled and explicit set of criteria on the basis of which the diagnostic status of homosexuality could be determined is evident from a symposium held on the subject at the annual APA conference of May 1973. The

[45] Although Rosenhan (1973, 252) acknowledges that such errors of commission (a "false positive" or type-II error) are prevalent throughout medicine in general, he argues that the stigma associated with mental disorder labeling imposes a responsibility upon psychiatrists to exert finer discrimination in diagnosis than their medical counterparts. Thus the issue Rosenhan raises is not so much about diagnostic reliability or validity – since after all, the pseudo-patients were recognized, within a relatively short period of time, as being symptom-free (Spitzer [1975]) – but rather about the merits of labeling.

symposium was organized by Robert Spitzer not only as a response to the highly-publicized protests that had dogged the conventions for the previous three years, but also in response to the growing opposition within the APA itself against the specific labeling of homosexuality as a mental disorder.

Participants	Reject or Retain Classification	Etiological component of definition
Robert J. Stoller, M.D.	Reject	Etiological
Judd Marmor, M.D.	Reject	Unspecified
Irving Bieber, M.D.	Retain	Etiological
Ronald Gold	Reject	Unspecified
Charles W. Socarides, M.D.	Retain	Etiological
Richard Green, M.D.	Reject	Consequentialist
Robert L. Spitzer, M.D.	Reject	Consequentialist

Table 2.1: Outline of positions on 1973 APA symposium on homosexuality. See accompanying text for details.

With respect to the appropriate set of criteria on the basis of which the diagnostic status of *any* putative mental disorder should be evaluated, a significant difference of opinion amongst the participants concerned whether *etiological* considerations are relevant to establishing that something is a mental disorder, or only the *consequences* of the condition. In other words, are the details of the biological or psychodynamic process that gave rise to the patient's condition relevant for deciding upon its psychiatric status? Or rather, does it suffice that the consequences of having the condition, for the afflicted person or his or her social group, are disturbing enough to warrant professional attention, regardless of its origin? Although the next four paragraphs will describe the content of

the debates, for the purpose of convenient reference, the following table (Table 2.1) provides a schematic outline that lists the participants, whether they retain or reject the classification, and whether or not etiological elements enter into their conception of "mental disorder" (or only consequentialist elements).

About half of the participants assume that etiological considerations are relevant for establishing the acceptability of a diagnostic category. Stoller, for example, argues that "homosexuality" is not a legitimate diagnosis because it does not satisfy the formal criteria for diagnosis: the existence of a syndrome and a uniform etiology (Stoller *et al*. [1973, 1207]). It is not a syndrome because there is only a single defining dominant feature (sexual preference) and empirical work suggests that there is no uniform etiological mechanism. Bieber, on the other hand, argues that the etiology of homosexuality is sufficiently uniform to warrant its inclusion. According to Bieber, something is a disorder if it involves deviation from a "biological program". He argues for the retention of "homosexuality" on the grounds that humans are "biologically programmed for heterosexual development", that the existence of this "program" is supported by empirical research on olfaction, and that this biologically programmed behavior is "dislocated" by early childhood pathological family dynamics (Ibid., 1210). Socarides also assumes that etiological considerations are relevant by arguing that something is a disorder if it represents deviation from a normal psycho-developmental trajectory. He argues for the retention of "homosexuality" by offering a psychodynamic theory that explains homosexuality as a failure to achieve full individuation from the mother in early infancy. He also claims that there are fairly successful treatments available for homosexuality, thus tacitly arguing that the existence of a treatment is a relevant criterion for inclusion (Ibid., 1212).

On the contrary, Green and Spitzer argue for the exclusion of "homosexuality" as a diagnostic entity on purely consequentialist grounds. Moreover, both explicitly argue that biological or psychodynamic etiology should be irrelevant to disorder classification. Green argues by stipulating a set of heuristics that should be employed in any decisions about classification. The strongest rationale for the inclusion of a putative disorder is that it evinces either "gross social dysfunction" or is associated with an inner emotional discord that "reduces the efficiency of behavioral functioning" (Ibid., 1213). Both criteria are indifferent to the mechanism by which the psychological condition is produced. He argues for the exclusion of "homosexuality" on the grounds that it is not intrinsically associated with either. He also argues on methodological grounds that psychodynamic theories of causation lend the weakest and most questionable support for the justification of disorder classification.

Spitzer's attempt to formulate a general definition of "mental disorder" formed the basis of his position on the diagnosis of homosexuality. According to the purely consequentialist definition he proposes there, a mental disorder must "either regularly cause subjective distress or regularly be associated with some generalized impairment in social effectiveness or functioning" (Ibid., 1215). On this basis, he argues that homosexuality *per se* cannot be judged disordered because it is not always associated with subjective distress or impairment in social functioning. The irony of this position, though, is that if someone *does* exhibit great subjective distress concerning his or her homosexuality, then, according to the definition, the person qualifies as having a mental disorder – namely, that of experiencing subjective distress about his or her homosexuality. Hence Spitzer's position was that reference to "homosexuality" should be dropped from the official nomenclature and replaced by "Sexual Orientation Disturbance". The latter category should be applied exclusively to homosexuals who are

distressed about their homosexual orientation and who seek help in resolving that distress.[46] One may ask, of course – as many psychiatrists did – why the distress that specifically concerns one's sexual orientation should fall under a distinct diagnostic category, rather than a broader and already existing category such as an anxiety disorder or phobic disorder. One may also ask why it should not apply more generally to anybody that exhibits distress concerning his or her sexual orientation and not merely to homosexuals.[47] One of the concerns shared by psychiatrists who opposed the original classification of homosexuality, such as Marmor and Green, was that the new diagnosis would simply reopen the threat of the psychiatric stigmatization of homosexuals.

Despite the acrimoniousness of the debates, however, there was a single thread of agreement that ran throughout them, which is that psychiatry should not contribute to the oppression, either legal or ethical, against homosexuals. Thus the crisis of legitimation was of ongoing concern for all of the participants. For those psychiatrists opposed to the diagnosis, this question of legitimacy became a weapon with which they attempted to undermine their opponents' arguments. Marmor, for example, raises this threat most forcefully by claiming that, "it is our task as psychiatrists to be healers of the distressed, not watchdogs of our social mores" (Ibid., 1209). Green, similarly, suggests that current classification reinforces "legal, religious, and other forms of social discrimination" (Ibid., 1214). In response, Bieber and Socarides argue that psychiatry has always opposed the oppression or stigmatization of homosexuals, and has done so precisely by removing

[46] A second round of debates within the APA, which lasted from March 1977 through February 1978, concerned the precise content and formulation of the new category (which became "Ego-Dystonic Homosexuality" in the DSM-III [APA (1980, 281)]). Unlike the first round of debates, this one was carefully guarded from being publicized, and was enacted through a large amount of internal memos and letters. (See Bayer and Spitzer [1982] for selected correspondence during this period.)

[47] Both of these alternatives to recognizing a distinct category for homosexuals were proposed by Green and Marmor in heated correspondence with Spitzer (Bayer and Spitzer [1982, 36-7]). Spitzer held fast to his original proposal, which eventually succeeded – but only for a short time. The category of "Ego-Dystonic Homosexuality", which appeared in the DSM-III of 1980, was dropped from the revised edition of the DSM-III, the DSM-III-R of 1987 (APA [1987]).

homosexuality from the realm of sin (moral realm) or the realm of criminality (legal realm) to the realm of disorder (medical realm). According to Bieber, Freud was the first to recognize that homosexuality is "a disorder of psychosexual development rather than as sinful and antisocial" (Ibid., 1211). Socarides echoes this sentiment in calling for the abolishment of legal discrimination against homosexuals: "It is unthinkable that homosexuals be persecuted for something over which they have no choice" (Ibid., 1213). Such considerations should help to undermine the simplistic conception that those psychiatrists who were in favor of retaining the classification should be seen as willing agents of social repression.

The APA Board of Trustees unanimously approved Spitzer's proposal in December 1973, to much publicity (Kirk and Kutchins [1992, 85]), and homosexuality was deleted as an independent diagnostic category (and replaced by "Sexual Orientation Disturbance") from subsequent printings of DSM-II.[48] It was widely seen as an effective "compromise position": it denies that homosexuality *per se* is a mental disorder, while at the same time it bestows a certain professional recognition and legitimacy upon those psychiatrists who wished to continue to offer therapy for distressed homosexuals, and who had, perhaps, made it into one of the cornerstones of their clinical practice. From the point of view of this chapter, what is important about Spitzer's definition is that it provides a way of resolving, to some extent, the legitimation crisis by offering a principled, albeit rudimentary, set of criteria for accepting or rejecting certain proposed psychiatric conditions – they must be associated with distress or impairment in functioning. At the same time, it represents a fairly atheoretical perspective on "mental disorder", and consequently, remains general enough to avoid interminable disagreement

[48] The struggle to delete "homosexuality" from the DSM-II was not completely over. In April of 1974, Socarides forced the Board of Trustees to submit their decision to a referendum of the APA membership; the APA voted by 58% to 37% to uphold the Board's decision (Kirk and Kutchins [1992, 88]).

amongst psychiatrists of different theoretical persuasions, thus deferring the problem of the disciplinary position of psychiatry.[49] On the basis of this achievement, Spitzer and other psychiatrists began the task of formulating a more precise and rigorous definition of "mental disorder" that could be used to help resolve other such disagreements as they arose.

2.1.2 The Task of Formulating and Approving a Definition of "Mental Disorder"

The 1973 debates directly motivated much of the later interest taken by members of the APA to explicitly define "mental disorder", and to use the definition in such a way that disagreements about the psychiatric status of an condition can be rationally resolved (Spitzer and Endicott [1978, 15]; Spitzer [1981, 211]; Klein [1978, 41]; Moore [1978, 87]). However, the task of formulating a definition of "mental disorder" (which had not been done in any official APA manual before 1980) had already been a concern of the Board of Trustees of the APA from the early part of 1973, prior to Spitzer's contribution to the symposium on homosexuality. The following will provide a historical overview of the events that led to the explicit definition of "mental disorder" that appeared in the DSM-III of 1980 and later editions of the DSM. In short, the following events took place over a seven-year period: the APA task force responsible for drafting the DSM-III was charged with the task of formulating and publishing a definition of "mental disorder"; among other proposals, Robert Spitzer and Jean Endicott presented, at an APA conference, a controversial definition of "mental disorder" according to which mental

[49] This fragile consensus did not last for long, although the definition satisfied the original political motivation for which it was formulated. Spitzer, in fact, was able to get some additional political leverage out of the definition before he abandoned it in favor of a new formulation. He invoked the definition in a letter, dated December 29, 1975, to the Committee of Black Psychiatrists to argue that, as against their proposal, racism does not qualify as a mental disorder. He writes, "As you know, we are still struggling with the problem of defining what is a mental disorder. With our current working definition racism would not meet the criteria for a mental disorder since it is only in certain environments that it is associated with distress" (Ibid., 102).

disorders form a subset of medical disorders; the attempt to subsume mental to medical disorders prompted harsh criticism; consequently, the task force voted against the inclusion of such a definition in the DSM-III, although they approved a modified definition according to which mental disorders stem from "internal dysfunctions" on the part of the individual.

Charge to Define "Mental Disorder"

In March of 1973, Walter Barton, then chief executive of the APA, sent a memo to the chairman of the APA Council on Research and Development informing him of the Board's instruction to appoint a task force to revise the DSM-II and prepare the DSM-III (Kirk and Kutchins [1992, 79]). In addition to recommending a more problem-oriented and quantitative diagnostic system, the memo also recommends, "the formation of a Task Force to Define Mental Illness and What Is a Psychiatrist", and that this definition be used as a preamble to the DSM-III (Ibid., 80).

In April 1974, the Committee for Research and Development reconstituted the Task Force on Nomenclature and Statistics (which oversaw the development of the DSM-III), and appointed Spitzer as its chair. Spitzer was, perhaps, an obvious candidate: he had already served as one of the core members of the DSM-II Task Force; he had written or coauthored several technical papers on diagnostic reliability; and his reputation for being politically astute was solidified due to the initiative he had taken to meet with gay rights protesters in 1972 and, on that basis, to organize the 1973 symposium by which the debates were, at least temporarily, resolved (Bayer and Spitzer [1985, 188]; [Wilson 1993, 404]).

As chair of the Task Force on Nomenclature and Statistics, Spitzer was able to select the core members of the new DSM-III Task Force, which consisted of five psychiatrists, two psychologists, and one biometrician (Millon [1986, 38]). The

psychiatrists of the Task Force were primarily research-oriented rather than clinically-oriented, and all of them endorsed the use of "operational" or criterion-based definitions for specific disorders which became recognized as one of the primary innovations of the DSM-III.[50] (See Section 2.3.2 on the philosophical significance of the use of operational definitions.)

The Task Force worked rapidly on the preliminary draft of the DSM-III, which was presented to the APA in May of 1975 (Ibid., 40). During the same month, the DSM-III Task Force also began to attempt a definition of "mental disorder" as per the recommendation of the Board of Trustees, and allowed any member of the Task Force to submit voluntarily his or her own proposal for approval (Ibid., 45). (The independent "Task Force to Define Mental Illness and What Is a Psychiatrist" recommended in 1973 by the Board of Trustees never materialized.)

Surprisingly, Spitzer at first did not volunteer the definition that he had attained so much political leverage from. This was due, at least in part, to inadequacies that he perceived in the original formulation and which he openly shared with other members of the APA in a memorandum dated May 28, 1976 (Bayer [1981, 169]). The major difficulty was not unrelated to the problem of contriving an adequate formulation of the diagnosis that was to replace "homosexuality". The problem was that there are several diagnoses that fall under the sexual deviations ("Paraphilias" in the DSM-III [APA (1980, 266)]) that do not typically result in "subjective distress" or "generalized

[50] This group is often said to have constituted the core of an "invisible college" (Blashfield [1984, 43]; Kirk and Kutchins, [1992, 97]); that is, the members had already been in frequent professional contact with one another, they were primarily located at a small number of research universities, and they had co-authored papers with one another or generously cited other members' papers in their own work. (The term "invisible college" itself alludes to the Invisible College of London and Oxford, the predecessor to the Royal Society of London.) Klerman (1978, 105) later referred to this group as the "Neo Kraeplineans" after Emil Kraeplin, who had largely initiated and advocated the differentiation and systematic classification of mental disorders on the basis of careful clinical and longitudinal observations. The prototype of criteria-based classification for mental disorders is Feighner *et al*. (1972). This prototype is developed and elaborated in Woodruff *et al*. (1974) and Spitzer *et al*. (1975).

impairment in social effectiveness or functioning", such as fetishism, transvestism, and voyeurism. Nonetheless, unlike homosexuality, there was a large consensus within the APA, and the mental health profession generally, that these conditions are undoubtedly mental disorders, regardless of whether they happen to be "ego-syntonic" for their bearers and regardless of the level of social adjustment the person may evince (Bayer and Spitzer [1982, 33]; Bayer [1981, 169]). Consequently, Spitzer believed that his "subjective distress or generalized impairment" criterion was not a necessary condition for having a "mental disorder", despite the fact that he presumably continued to believe it to be sufficient.

When Spitzer eventually returned to the task of formulating a proper definition of "mental disorder" (see below), he attempted to circumvent this problem by introducing a new disjunct to the "distress or disability" clause, namely, the notion of an "inherent disadvantage" (Bayer [1981, 169]). Spitzer adopted the notion of "inherent disadvantage" from Kendell (1975b) whose concept of "biological disadvantage" will be described in Section 2.3.1. Briefly, according to Spitzer, a psychological condition places its bearer at an "inherent disadvantage" if it prohibits the person from satisfying "important psychological or biological needs" (that is, it is "disadvantageous"), and that it does so in almost every environment (that is, it is "inherent" to the condition rather than relative to a specific environment) (Spitzer [1981, 212]; see Spitzer and Endicott [1978, 19-23] for an extended discussion of this condition). One important such psychological need is served by the "ability to experience sexual pleasure in an interpersonal context" (Spitzer and Endicott [1978, 26]). Homosexuality, of course, is not prohibitive in this respect. However, many of the paraphilias, such as fetishism and zoophilia, are prohibitive in this respect, and "inherently" so – that is, they are prohibitive in almost every conceivable environment (Bayer [1981, 169]) to the extent that they do not involve other persons.

It hardly need be pointed out that this additional criterion leaves the selection of what constitutes an "important biological or psychological need" completely open to the interpretation of the individual psychiatrist (Bayer [1981, 169]; Spitzer and Endicott [1978, 32]). Spitzer emphasizes this point in a later paper (Spitzer [1981, 212]), in which he acknowledges that it constitutes a value judgement on the part of the psychiatrist that often cannot be justified. He also notes that this sort of *ad hoc* theoretical maneuver gave rise to the suspicion among some psychiatrists that the whole project of contriving a definition of "mental disorder" would be irrelevant to actual decision–making protocols; instead, they believed, decisions would be made, "and then the definition tinkered with to justify them" (Spitzer [1978, 16]).

The first member of the DSM-III Task Force to volunteer a position paper on the definition of "mental illness" was Donald Klein (Millon [1986, 44]). Although Klein's initial draft is unpublished, it will be assumed that it is not substantially different from a version published in 1978 (Klein [1978]) in an anthology jointly edited by Klein and Spitzer. The concept of "illness" presented by Klein has two components. First, there must be an "involuntary impairment" of an *evolved function*; second, this impairment must be of sufficient magnitude to elicit the appellation of the "sick role" by the individual's social group (Ibid., 46). This definition will be elaborated and evaluated in Section 2.3.3. What is important, in this context, is the medical orientation of the definition. Klein proposes the notion of an "involuntary impairment of an evolved function" as a way of conceptualizing mental and medical disorders as having the same objective basis: according to Klein, mental illness is merely a subset of illness in general that "presents evidence in the cognitive, behavioral, affective, and motivational aspects of organismal functioning" (Ibid., 70), rather than in the physiological aspects of organismal functioning. Moreover, a "dysfunctional state" is explicitly defined there in terms of its

etiology, and not its consequences: it consists in "a suboptimal deviation from [an] evolutionarily determined process and the result of disease" (Ibid., 51).

Klein's proposal was unsatisfactory to the Task Force, which perceived it as "overly abstruse and theoretical" (Millon [1986, 44]). Spitzer, in particular, was worried that the definition would be accepted by default (in the absence of alternative proposals) and hence he, along with Jean Endicott, set to work on a new formulation – a formulation that nonetheless incorporates Klein's emphasis on disorder as deviation from "organismic functioning". The Spitzer-Endicott definition was first presented to the APA in May 1976 and a version was published in 1977 (Spitzer *et al*. [1977]).[51] A significantly modified version of this definition ultimately made it to publication in the DSM-III, in the introduction which was authored by Spitzer himself.

The Spitzer-Endicott Definition of "Mental Disorder" and the "Medical Model" of Psychiatry

The basic elements of the definition that Spitzer had presented in 1973 – that a mental disorder must be associated with subjective distress or disability – were retained and considerably amplified in the 1976 APA presentation and the 1977 article. According to their definition,

> [Mental disorders,] in their extreme or fully developed form…are directly associated with either distress, disability, or, in the absence of either of these, disadvantage in coping with unavoidable aspects of the environment. Furthermore, they are not quickly ameliorated by simple nontechnical environmental maneuvers or informative procedures and do not have widespread social support. (Spitzer *et al*. [1977, 4])

[51] A more elaborate version was published in 1978 (Spitzer and Endicott [1978]), some features of which will be critically discussed in Section 2.3.2.

Historically, however, the most controversial feature of this article is not the definition *per se* but the claim that mental disorders are a "subset of medical disorders":

> These principles [i.e., criteria for defining "mental disorder"] help to avoid an overly broad definition of mental disorders that would view all individual and social unrest or problems of living as psychiatric illness, and at the same time justify the designation of mental disorders as a subset of medical disorders. (Ibid.)

Note that the justification offered here is a direct response to the two crises described earlier, the legitimation crisis and the crisis of the disciplinary position of psychiatry. On the one hand, the concern to avoid an "overly broad" definition of mental disorder is a reflection of the legitimation crisis, namely, that psychiatry was unable to distinguish disorders from "all individual and social unrest" and "problems of living". (The expression "problems in living" was made famous by Szasz [1961], who popularized the notion that what psychiatrists refer to as "mental disorders" are, instead, rather commonplace "problems in living". Hence the passage above is a tacit response to the antipsychiatric tradition.) On the other hand, the concern to subsume mental disorders under medical disorders is a way of affirming the medical orientation of psychiatry and thereby attempting to distinguish psychiatry from the other mental health professions; hence it constitutes a response to the problem of clarifying psychiatry's disciplinary position.

The attempt to subsume, at least conceptually, mental to medical disorders provokes the question of what a "medical disorder" is. Although Spitzer *et al*. do not in this context provide a definition of "medical disorder", they do define what they call the "medical model" of psychiatry, which they believe psychiatrists ought to endorse. According to Spitzer *et al*., the "medical model" is simply the hypothesis that "there are organismic dysfunctions which are relatively distinct with regard to clinical features,

etiology, and course" (Ibid., 5). Hence, being the same as, or at least produced by, an "organismic dysfunction" is a necessary condition for being a medical disorder and, therefore, if the proposed subsumption holds, for being a mental disorder as well.

A year later, in a revised and slightly more sophisticated version of the article, Spitzer and Endicott (1978) work out this way of conceptualizing mental disorders more carefully. According to them, a mental disorder is a type of medical disorder the sole differentia of which is that the organismic dysfunction that produces it is primarily manifested psychologically or behaviorally rather than physiologically. In this respect it is, in spirit, almost identical to Klein's (1978) definition (described above). According to Spitzer and Endicott:

> A medical disorder is a relatively distinct condition resulting from an organismic dysfunction which in its fully developed or extreme form is directly and intrinsically associated with distress, disability, or certain other types of disadvantage. The disadvantage may be of a physical, perceptual, sexual, or interpersonal nature. Implicitly there is a call for action on the part of the person who has the condition, the medical or its allied professions, and society. (Spitzer and Endicott [1978, 18])

Note that this definition of "medical disorder" incorporates Spitzer's "distress or disability" criterion from the 1973 APA symposium, with the addition of the concept of "disadvantage". Additionally, it stipulates that the distressing, disabling, or disadvantageous condition results from an "organismic dysfunction". On the basis of this definition of "medical disorder", they define "mental disorder" fairly straightforwardly: "A mental disorder is a medical disorder whose manifestations are primarily signs or symptoms of a psychological (behavioral) nature, or if physical, can be understood only using psychological concepts" (Ibid.).

Rejection of the Proposed Definition

As noted above, Spitzer and Endicott first presented a version of their definition at the APA conference of 1976. This presentation provoked heated criticism, predominantly by psychologists affiliated with the American Psychological Association, and primarily with respect to the proposed subsumption of mental disorders to medical disorders (Millon [1983, 806]). For example, in an article entitled "But is it Good for the Psychologists?", Schacht and Nathan (1977) argue that Spitzer and Endicott's appeal to "organismic dysfunctions" illegitimately abstracts from the environment by placing the disorder solely "within" the organism (Ibid., 1023). They also claim that the primary function of the locution is to expand the professional jurisdiction of psychiatry and diminish that of the other mental health professions (Ibid., 1024). (Garmezy [1978, 6] levels a similar charge of "territoriality" against the DSM-III as a whole.) Zubin (1977) argues that the most serious flaw with the DSM-III is the "false or at least unproven", and "entirely gratuitous", assumption that mental disorders are medical disorders (Ibid., 6). McReynolds (1979) writes that the "disease conception of behavioral disturbance" advocated by Spitzer and Endicott is an example of a once-useful heuristic which "now entraps our thinking and limits our research and practice" (Ibid., 125).

The ensuing correspondence that took place between Theodore Blau, then-president of the American Psychological Association, and Jack Weinberg, president of the APA, reveals the extent of the tension between the organizations that erupted as a result of the 1976 presentation and the ensuing 1977 publication of the definition. In a letter dated August 8, 1977, Blau expressed his concern with the possibility that a definition of "mental disorder" that subsumes mental to medical disorders would be published within the DSM-III. In it, he points out that, "[o]f the 17 major diagnostic classes, at least 10 have no known organic etiology" (cited in Kirk and Kutchins [1992,

112]). Upon receipt, Weinberg passed the letter on to Spitzer and asked him to draft a response. Spitzer's draft was openly combative, and Weinberg largely maintained the combative tone in his response to Blau, dated November 3, 1977. After some conciliatory remarks to the effect that he would consider publishing a disclaimer along with any proposed definition of mental disorder, Weinberg writes:

> Where are we to go from here? You can continue to try to convince us that most mental disorders in the DSM-III classification are not medical disorders. You will not only fail to convince us, but we believe that it is inappropriate for you to attempt to tell us how we should conceptualize our area of professional responsibility. You can try to convince us that even if we believe that mental disorders are medical disorders, we should not explicitly say so in DSM-III. You will not convince us of this either. We believe that it is essential that we clarify to anyone who may be in doubt, that we regard psychiatry as a specialty of medicine. (cited in Ibid., 114)

Blau responded by arguing that the APA's attempt to carve out its area of professional responsibility seems to exclude the contribution of the other mental health services, and, moreover, that it "suggests disdain" for those services: "Using the concepts of 'mental' and 'medical' synonymously or inclusively may exclude or deny the promising independent research and service [of those professions]" (cited in Ibid.). Blau continues with a more general criticism of the whole project of drafting the DSM-III:

> Candidly DSM-III, as we have seen it in its last draft, is more of a political position paper for the American Psychiatric Association than a scientifically-based classification system. To continue to promulgate a classification system that does not meet the needs of emotionally troubled persons is not in the best interest of society or of either of our professions. (cited in Ibid., 115)

The attempt to incorporate a definition of "mental disorder" within the DSM-III was ultimately rejected by a majority vote by the DSM-III Task Force in February of

1978. Part of the reason it was rejected was because a liaison committee consisting of three psychologists associated with the American Psychological Association had been given voting power. Presumably, they shared the critical attitude evinced by the American Psychological Association towards the definition (Millon [1983, 806]). In April of the following year, however, the Task Force approved a modified version of the definition to be inserted in the glossary of the DSM-III (Ibid., 806). Although the version that made it to publication in 1980 is not there proposed as a "definition" of "mental disorder", it is proposed as a working characterization that is similar to the earlier Spitzer-Endicott definition. (A slightly more elaborate form of the characterization is presented in the introduction of the DSM-III itself.)

Given the Task Force decision, the offending passage that mental disorders are a "subset of medical disorders" does not appear. It does, however, specify that a mental disorder must stem from a "dysfunction" on the part of the individual – thus invoking the conceptual apparatus of the so-called "medical model" described by Spitzer *et al*. (1977) in their contentious paper. The notion of "dysfunction", however, is left undefined. This problem will be returned to in Section 2.2.2, after examining the DSM-III characterization itself.

2.2 "MENTAL DISORDER" IN THE DSM-III

Spitzer's introduction to the DSM-III explicitly renounces the attempt to provide a definition of "mental disorder": "…there is no satisfactory definition that specifies precise boundaries for the concept 'mental disorder'" (APA [1980, 6]). Nonetheless, it points out that, "it is useful to present concepts that have influenced the decision to include certain conditions in DSM-III as mental disorders and to exclude others" (Ibid.).

Hence the following should be viewed as a working characterization rather than, strictly speaking, a definition of "mental disorder":[52]

> In DSM-III each of the mental disorders is conceptualized as a clinically significant behavioral or psychological syndrome or pattern that occurs in an individual and that is typically associated with either a painful symptom (distress) or impairment in one or more important areas of functioning (disability). In addition, there is an inference that there is a behavioral, psychological, or biological dysfunction, and that the disturbance is not only in the relationship between the individual and society. (When the disturbance is *limited* to a conflict between the individual and society, this may represent social deviance, which may or may not be commendable, but is not by itself a mental disorder.) (Ibid.)

The characterization provided in the DSM-III-R (APA [1987]) contains some minor modifications of the original. It specifies that the syndrome or pattern must "currently be considered a manifestation of a behavioral, psychological, or biological dysfunction *in the person*" (Ibid., xxii; emphasis added).[53] In addition to excluding conflicts between the individual and the society, it also specifies that the syndrome must neither be "merely an expectable response to a particular event, e.g., the death of a loved one" (Ibid.), nor deviant behavior, "e.g., political, religious, or sexual" (Ibid.).[54]

There are two points of interest about this characterization that are of central importance for motivating the explication of "function" that will be presented in the following chapters. The first concerns the prominent use of the notion of a dysfunction on

---

[52] Interestingly, unlike DSM-III, the characterization offered in the revised version of DSM-III, DSM-III-R, *is* explicitly offered as a "definition" of the concept (APA [1987, xxii]).

[53] Presumably, this definition would exclude relationship disorders, for example, marital or family disorders in which the dysfunction cannot be localized to one or the other member, but describes maladaptive patterns of interaction. Widiger and Trull (1991, 115) criticize this restriction, remarking that, "the validation of interpersonal and systems models of pathology [is]…hindered by a taxonomy that recognizes only organismic dysfunction". However, Eysenck *et al*. (1983) question the usefulness of assessing disorders in units larger than the individual.

[54] The 10th edition of the *International Classification of Diseases* (ICD-10), published by the World Health Organization (WHO), includes a special mental and behavioral disorder classification (WHO [1992]). Published in 1992, it was developed in collaboration with the APA and includes a similar characterization (WHO [1992, 5]).

the part of the person to establish a contrast with the case of political, religious, or sexual deviance, or more generally, (mere) "conflict between the individual and society". This suggests that appeal to the notion of an internal dysfunction is meant to provide an explication of the idea that, when an individual has a mental disorder, there is a sense in which *something has gone wrong within that individual*, and thus that the ascription does not merely reflect the judgement that the person in question acts, feels, or thinks in ways that do not satisfactorily conform to social expectations and values. Spitzer and Williams (1982) make this intuition explicit: "The assumption that something has gone wrong with the organism…is in the DSM-III expressed in the phrase, 'there is an inference that there is a behavioral, psychological, or biological dysfunction'" (Ibid., 21). Indeed, one would expect the authors of the definition to express such a concern with separating mental disorders from mere conflicts with social norms, given the legitimation crisis that motivated the attempt to formulate the definition in the first place. This raises a problem, however, since to oppose "dysfunction" and "deviance" is only to provide a negative characterization of "dysfunction"; it does not contribute to providing a positive characterization of what constitutes an internal dysfunction *as such* (Wakefield and First [2003, 35]).

The second point of interest about this characterization is the apparently ambiguous appeal to the notion of "functioning". On the one hand, the first sentence of the DSM-III characterization specifies that the syndrome or pattern is associated with distress or disability, where "disability" is defined as "impairment in one or more important areas of functioning". On the other hand – and this is supposed to amplify the first sentence – the second sentence states that there must also be a "behavioral, psychological, or biological dysfunction". Consequently, *disability* as "impairment in an area of functioning" refers to something other than a "behavioral, psychological, or

biological" *dysfunction*. The revised definition (of the DSM-III-R) clarifies, if not the meaning of these terms, that the dysfunction must be "*in* the person" and that the distress or disability must be a "manifestation" of the dysfunction. Hence the DSM-III alludes to *two* different types of "functioning", and suggests that the relation between the two is a causal one: the dysfunction "in the person" causes the "manifest" impairment in an area of functioning ("disability").

In order to understand this characterization of "mental disorder", and in particular, the concepts of functioning that it embodies, the following two questions should be answered: first, how should the notion of "disability" be explicated?; second, how should the notion of "dysfunction" be explicated? (It will be assumed, for the time being, that the concept of "distress", while not entirely unproblematic, is conceptually clear enough that it does not warrant further consideration here.)

2.2.1 "Disability" as Failure of Social-Role Functioning

The notion of "disability" is defined in DSM-III as "impairment in one or more areas of functioning". This gives rise to the question of what constitutes an "area of functioning". Fortunately, this question is not difficult to answer, since the specific "areas of functioning" that must be affected in order for something to qualify as a mental disorder are typically listed separately for each specific diagnostic category, and are either embedded within the formal diagnostic criteria themselves, or in the associated, informal description. For example, in the DSM-III, a criterion for Schizophrenic Disorder is "deterioration in functioning in such areas as work, social relations, and self-care" (APA [1980, 189]). Attention Deficit Disorder with Hyperactivity is typically associated with impairment in "academic functioning" (Ibid., 42); Avoidant Disorder (labeled "Shyness Disorder" in an early draft; see Garmezy [1977, 4]) with "social functioning in peer relationships" (Ibid., 55). Typically, however, such descriptions are limited to

stipulating some degree of impairment in the rather generic category of "social and occupational functioning". Spitzer introduced these "clinical significance criteria" in order to remind the clinician to evaluate whether the presenting symptoms are sufficiently severe to warrant diagnosis, although he currently doubts their clinical utility (Spitzer and Wakefield [1999, 1863]).

What these listings show is that notion of an "area of functioning" is not conceived, as it standardly is in the biological sciences, in terms of the function of a trait, or a vital function, such as respiration, metabolism, reproduction, and so on – those very general organismic capacities in the absence of which the organism's survival or reproductive ability is impaired. Rather, it is conceived in the sense of a failure to carry out certain roles that largely define a person's contribution to a social group, or that constitute the *sine qua non* of such group participation, such as being a co-worker, a friend, a peer, a student, or more generally, being "presentable" to others in the sense of maintaining an acceptable level of personal hygiene, dress, and so on (Ibid., 1858). It is essentially a relational criterion in that it draws attention to one's "place" within a greater social structure.

The use of the term "disability", however, to gloss the notion of "impairment in an area of functioning" is misleading, because the notion of a "disability" typically implies an *involuntary* inability to engage in a certain activity. However, the DSM-III characterization does not specify whether the impairment is voluntary or involuntary. A person who typically does not remain employed for long periods of time, does poorly academically, or does not tend to sustain long-term intimate relationships, is not necessarily thought to have a "disability". The person may simply be uninterested in maintaining long-term employment, may lack interest in academic achievement, or may enjoy variety in his or her romantic relationships. Nonetheless, such a person may

correctly qualify as "functioning poorly" in the area of employment or marriage, in the sense that the person's behavior is seen as falling below the widely-shared standards or expectations of the community with respect to those institutions.

Consequently, in the following, the notion of "disability" that appears in the DSM-III should be understood as a *failure of social-role functioning*; in other words, it designates *unsatisfactory* performance with respect to these areas of functioning, where "unsatisfactory" refers to the norms of appropriateness that originate within, and possess whatever justification they may have by virtue of, the needs and goals of the individual's group. (This is not, of course, to say that the individual in question does not share those goals, or that the individual's inability to function properly is not a significant source of personal distress.) Thus it will be considered to be a sociological category. This is in keeping with the sociological orientation of Klein's (1978) definition of "mental illness", and specifically, with his notion that the concept of mental disorder implicitly prescribes the "sick role" (along with the waiver of certain rights or the assumption of certain privileges that attends to that role), due to a person's failure to satisfy the responsibilities associated with his or her normal social role.

The interpretation of "disability" as "failure of social-role functioning" implies that what the DSM-III characterizes as "social deviance", or "political, religious, or sexual" deviance, is itself a form of "disability" – since social deviance is, by definition, deviation from expected or valued social roles and responsibilities. Thus, appeal to the "distress, disability, or disadvantage" criterion alone is by no means sufficient for defining "mental disorder", since it is overinclusive – all forms of social deviance may fall under "disability". This raises the problem of what *differentiates* mental disorder, properly speaking, from all such forms of role failure, since the whole project of legitimizing the authority invested in psychiatry, and clarifying its disciplinary position in

a relatively uncontroversial manner, rests upon such a principled differentiation. This is presumably what the concept of an internal "dysfunction" is supposed to accomplish.

2.2.2 Presence of an Internal "Dysfunction" and the Amplification Problem

What is it to have an internal dysfunction? Unlike the concept of "disability" as failure of social-role functioning, this second criterion is suppose to provide a foundation for mental disorder classification that is not relative to the expectations and values of a person's current social environment, and to secure the medical credentials of psychiatry. It is surprising, then, that it remains undefined in the DSM-III, given its conceptual import. (Wakefield and First [2003, 35] also note that the most serious problem with the DSM-III definition is that "there is no explanation or analysis of the critical concept of dysfunction".) By leaving the term undefined, the authors of the DSM-III, and later editions, open themselves to the charge that the term does not *in fact* amplify the definition of "mental disorder", but rather, merely serves to reify the socially-relative nature of (at least some types of) norm-violating behavior by construing it as a non-relational property of the individual – that is, as an inner mechanism that disposes that individual to violate social norms and that falls within the province of medicine. In other words, the problem concerns whether the term "dysfunction" is merely a slogan that serves to obscure the two crises rather than to offer a substantive solution to them. Kendell (1986), whose attempt at a definition of "disease" will be discussed in Section 2.3.1, raises the same issue in arguing that the DSM's usage of "dysfunction" does not resolve any fundamental problems, but rather, is "vaguely worded [enough] to allow any term with medical connotations to be either included or excluded in conformity with contemporary medical opinion" (Ibid., 41).

Thus one is faced with what will be referred to as the *amplification problem*: whether, and how, the concept of an internal "dysfunction" amplifies the definition of

"mental disorder". Solving the problem requires either an explicit definition of "dysfunction" or a set of procedures that allow for its reliable determination and by which it could be implicitly defined. The amplification problem, then, serves to motivate the explication of the concept of an internal "function" and "dysfunction" that will occupy the next two chapters of the dissertation.

Adequacy Conditions on the Definition of "Function"

Part of the method that will be used to explicate "dysfunction" in the following chapters will involve the evaluation of various definitions that have been proposed in the psychiatric and philosophical literature. In Section 2.3, the proposals that will be evaluated are those that psychiatrists have offered; in Chapter 3, they will be primarily those of philosophers. However, before presenting and evaluating various such attempts, it will be helpful to set forth some of the demands that should be placed upon any definition of "function" that is appropriate to the context of psychiatry. These demands will take the form of two "adequacy conditions", that is, rules that inform one as to when a proposed definition even qualifies for further consideration. The justification for these adequacy conditions is not absolute and historically invariant, but rather, stems from the recent social and historical context of American psychiatry itself and the specific problems it has confronted as a discipline. To recapitulate, these problems involve the attempt to justify the social and political power invested within it (the legitimation crisis) as well as the attempt to clarify its area of professional responsibility (the disciplinary position crisis). American psychiatry – or at least those prominent representatives of the APA upon whom the profession bestowed a disproportionate share of the responsibility for articulating its self-conception – attempted to respond to these problems by drawing attention to the medical orientation of its practitioners and, on this basis, appealing to the

notion of an internal or organismic "dysfunction" to define partly that area of expertise. The undefined status of the term gives rise to the amplification problem.

Although the crucial *definiendum* in the following is, of course, "dysfunction" (or "malfunction", etc.) the adequacy conditions presented here will impose constraints upon the definition of "function" instead. One *a priori* justification for this stems from the intuition that *being dysfunctional* represents some type of privation or aberration of something's capacity to perform its normal function and that this dependency relation should be reflected in the definition of those terms – the consequence being that "dysfunction" should be defined in terms of "function" and not the other way around. A more pragmatic justification for this is that most of the relevant conceptual analyses offered by philosophers for defining "dysfunction" – many of which will be reviewed in Chapter 3 – follow the same procedure. Because of this, by imposing adequacy conditions upon "function" rather than "dysfunction" one can appropriately engage with that burgeoning literature.

The first condition of adequacy (CA) on any definition of "function" that is appropriate for psychiatry is relatively trivial, given the motivation for its introduction:

> CA_1: If *X* has a function – where *X* is some organ, trait, or part – then *X* is capable of being dysfunctional.

In other words, no concept of function that does not permit the explication and applicability of a corresponding concept of "dysfunctional" and not merely "non-functional" can satisfy the conceptual demands that are imposed upon it. Despite its obviousness in this context, this condition of adequacy must be stated explicitly, since some prominent explications of "function" do not satisfy this condition. In order to satisfy this condition, a definition of "function" must allow a conceptual distinction to be

drawn between *having* a function and *performing* a function, since presumably, in order for a trait to be "dysfunctional", it must possess a function that it does not perform. For example, one biological definition of "function" as "all physical and chemical properties arising from [organismic] form [with the exception of those that refer to the organism's environment]" (Bock and von Wahlert [1965, 274]) does not allow this distinction to be drawn and hence does not satisfy CA_1, since according to this definition, virtually any activity of a trait qualifies as its "function". However, as will be seen, merely allowing this distinction to be drawn is not sufficient for defining "dysfunction" (see Section 2.3.3). For example, something may actively prohibit something else from performing its function, and there are good reasons not to say that the latter thing is "dysfunctional" or "malfunctioning", even though it has a function that it cannot perform.

The second adequacy condition is an extension of the first in the sense that it places constraints upon the manner in which something can be "dysfunctional". This adequacy condition, however, is less obvious than the first, so some motivation is necessary. Roughly, the idea that motivates the second adequacy condition is that whether or not a trait is "dysfunctional" should not depend upon whether someone happens to value, or disvalue, the activity of the trait. If this were so then the "dysfunction" criterion would not necessarily amplify the definition of "mental disorder", but merely serve to introduce a value judgement concerning the distressing, disabling, or disadvantageous condition, in lieu of a designation for the inner cause of the disturbance. For this reason, a preliminary – albeit inadequate – attempt to formulate the adequacy condition may be the following:

> CA_2: Whether or not *X* is dysfunctional is not determined by changes in the way that somebody either values *X* or values the effects that *X* produces.

For example, if there is a gene that disposes one to homosexual orientation, then whether or not the gene's activity is "dysfunctional" should not depend merely upon changes in moral attitudes about sexual orientation. One of the consequences of CA_2 is that the concept of "dysfunction" should not be an *evaluative* term in the sense of Hare (see Section 1.3).

However, this formulation introduces an unnecessary restriction, because it does not allow for the possibility that a trait could become dysfunctional precisely because someone negatively values its activity, and as a consequence, attempts to suppress the activity, and thereby interferes with the trait's normal functioning, and makes it dysfunctional. So, for example, suppose that homosexuality is conceived as a psychological disposition one of the functions of which is to bring about emotional and sexual fulfillment with another person. Suppose, then, that many people come to disvalue it, and because they disvalue it, they attempt to inhibit its expression. As a consequence, many homosexuals develop certain psychological conflicts about their sexual orientation that lead to distress rather than emotional and sexual fulfillment. Insofar as the same disposition now produces distress, it may perhaps be said to qualify as "psychologically dysfunctional". In fact, the authors of the DSM-III recognize the possibility that "ego-dystonic homosexuality" may have such an etiology: "The factors that predispose to Ego-dystonic Homosexuality are those negative societal attitudes toward homosexuality that have been internalized" (APA [1980, 282]). Hence the second adequacy condition should not imply that whether or not something is dysfunctional must be *causally* independent of social values.[55]

---

[55] More formally, the type of situation that should not be excluded on *a priori* grounds from causing an inner "dysfunction" is the following: A person, *P*, has an inner disposition, *D* (e.g., homosexual orientation), that give rise to behavior *B* (homosexual behavior); *B* is disvalued by a person, *Q* (where *P* and *Q* may be the same person) and, as a consequence of this negative evaluation of *B*, *Q* interacts with *P* and creates a new disposition, *D′* (anxiety about homosexual orientation) within *P*; the interaction of *D* and *D′* qualifies as a "dysfunctional" interaction or an internal "dysfunction". Moreover, one should not

The problem with CA_2, then, is that it does not specify that changes in the functional status of a trait's activity should not be brought about *merely* by changes in the way that someone values that activity, if such changes do not in have any effect upon the characteristic structure or activity of that trait. This lack of specificity can be resolved by a revision of CA_2 that takes this causal independence into account (although this, too, will be abandoned in favor of a different formulation):

> CA_2': Whether or not *X* is dysfunctional is not determined by changes in the way that somebody either values *X* or values the effects that *X* produces when such changes have no effect upon the characteristic structure and activity of *X*.

The simple justification for CA_2' is that if the notion of "dysfunction" in this context is to explicate the intuitive idea that something has *gone wrong within an individual*, then in order for something in the individual to become dysfunctional, something *within that individual* must be affected and not merely something within, say, the minds of those within that individual's group. One of the situations that CA_2 excludes, and that CA_2' permits, then, is that homosexuality might be a manifestation of a psychological dysfunction in very homophobic societies, and not in more tolerant societies, as a consequence of the way in which negative social attitudes about homosexuality are internalized by a person.

However, CA_2', while adequate to the intuition that motivates it, is unnecessarily narrow given the rationale that justifies it. For this rationale supports a much stronger adequacy condition, namely, that *any* change – and not merely one related to social attitudes – that has no effect on *X* should not be able to determine whether or not *X* is

---

immediately exclude the conceptual possibility that if the interaction between the components of a system is dysfunctional, then the components themselves should be thought of as "dysfunctional". If this is the case, then *D* in *P* is now dysfunctional, and if *D* continues to give rise to *B*, then *B* is now caused by an internal dysfunction as a result of *Q*'s negative evaluation of *B*.

dysfunctional. In other words, what should be recognized as the core intuition regarding the psychiatric concept of "dysfunction" is not that its ascription should not be relative to the values of a social group, but that it must not be "externalistically individuated" – that is, the difference between the case in which something within an individual is malfunctioning, versus that in which it is not, should not be determined exclusively by changes that take place outside of that individual. Rather, it should supervene upon changes within the individual himself or herself.

The concept of "externalism" (as opposed to "internalism") originated with the theory of language associated with Putnam (1975) and Kripke (1972), according to which the meanings of many of the words that one uses – particularly those of proper names and so-called "natural kind" terms – are not determined by the psychological state of the speaker (e.g., "what one means by them") but rather, by features of the social environment or natural world that they refer to. This view was elaborated into a theory of mind by McGinn (1977) and Burge (1979), according to which the content of one's mental states are determined by features of a person's social or natural environment, regardless of whether or not those features have any actual effect upon the intrinsic characteristics of the person's experience, such as the qualitative features of that experience, or upon the structure of the person's brain or nervous system. A consequence of externalism in the philosophy of mind is that, were a person to be transposed from one environment to another, the contents of that person's mental states could change as a function of that transposition, even in the absence of any experiential change, or of any changes in the world that physically affect that person. The situation is analogous to driving a car at a constant speed down a long road with no signs, but where the speed limit continues to change – at one moment, one is violating the law; at another, one is not; although one's beliefs and actions can be described as unchanging with respect to the

law. It is this sort of externalism that, although appropriate, perhaps, for some aspects of language, mind, and the legal status of actions, should not determine whether or not a trait is dysfunctional. Rather, for a trait to become dysfunctional, some change must occur that affects the characteristic structure or activity of the trait itself.[56]

Of course, the question of what qualifies as an "internal", rather than an "external", feature of a person probably does not admit of any precise answer. For example, does "internal" simply mean, "beneath, and inclusive of, the skin"? But what constitutes the "inside" of the body is typically a function of the type of model that one is using to analyze it: an immunological model, for example, will consider many items that are "under the skin" to nonetheless qualify as "foreign" to the organism. More importantly, a person's experience is not necessarily physically "localizable" in the same way that a person's organs are, and hence it may be less clear how to draw a precise boundary between the "internal" and the "external" in those cases. Finally, there is a sense in which, in the case of behavior, the distinction may not be intelligible. For example, the category of "job-seeking behavior" relies so heavily not merely upon what a person takes himself or herself to be doing, but what other people understand that person to be doing, that changes in the latter may altogether change the type of behavior in question.

Nonetheless, in most cases it will be fairly clear whether the model that is used to differentiate the dysfunctional from the non-dysfunctional activity of a trait appeals only to "external" features of the individual, or "internal" ones as well. For example, a model of depression that attributes the dysfunction to lowered serotonin production qualifies as

[56] There is perhaps a trivial sense in which *no* definition of "function" can satisfy this criterion – namely, insofar as *every* physical change in an individual's environment will presumably bring about *some* change, however minimal, in the physical constitution of the individual, such as a nearby leaf's disrupting the air molecules that affect one's skin. Therefore, this discussion presupposes that there are some constraints upon which sorts of environmental changes suffice to change the "characteristic structure or activity" of a trait, and which do not, although these constraints will largely by left implicit.

an "internal" model; a model of simple phobia that attributes the dysfunction to a repressed, unconscious trauma also qualifies as an "internal" model; but a model of schizophrenia that attributes the dysfunction merely to large-scale changes in social arrangements, with no implication that such changes interact with the psychological or biological profile of those individuals who come to have schizophrenia, would qualify as an "external" model. In this context, then, it is not so important to specify rigorously a precise boundary between "internal" and "external" changes to a trait as much as it is to secure the intelligibility and general applicability of the distinction in the first place.

Within this consideration in mind, CA_2' will be generalized as follows:

> CA_2*: Whether or not *X* is dysfunctional is not determined by any changes that have no effect upon the characteristic structure or activity of *X*.

This generalized formulation has two immediate advantages over its predecessor. The first, of course, is that CA_2' follows from it as a special case. The second is that it supports the intuition that whether or not a trait is dysfunctional should not merely be determined by whether its activity deviates from a *statistical* norm, for example, from the average value of that trait within a population or by the relative frequency of the trait. For example, according to Kendell (1975b), Lord Cohen defined illness as a "deviation from the normal…by way of excess or defect" (cited in Kendell [Ibid., 309]); Kendell rightly points out that the definition is inadequate because it does not distinguish between those deviations that are harmful, those that are neutral, and those that are beneficial to their bearers. Correspondingly, CA_2* does not permit mere statistical deviations from normalcy from counting as "dysfunctional", because one can arbitrarily make the bearers of a given trait exceedingly rare, or exceedingly common, depending on the sorts of changes one induces in the bearers of variant traits within the population, where the

bearers of the one type of trait are isolated from the bearers of the variant traits. This implies that although the relative fitness of having that trait can change as a function of such purely environmental changes, a trait that is at one time functional cannot be rendered dysfunctional by means of such changes. Rather, in order for a functioning trait to become dysfunctional, something in the environment would have to interact with it in such a way as to actually change its characteristic structure or activity.

One might argue that CA_2* is *too* broad, because it introduces a conceptual distinction between the relative fitness of a trait in its current environment and the functional status of that trait. Yet according to one prominent analysis (Bigelow and Pargetter [1987]), a trait has a "function" if it is disposed to contribute to the survival or reproduction of the organism that possesses it, in that organism's natural habitat. (See Section 3.1.3., under "Fitness-Contribution Theories".) Now, suppose that one were to use this definition to construct a corresponding concept of "dysfunction" – for example, a trait is "dysfunctional" if it is disposed to *reduce* the fitness of the organism possessing it. Suppose, furthermore, that one analyses the concept of "disposition" simply in terms of relative frequency: for example, if one were to say that trait *T* is disposed to reduce fitness if bearers of *T*, more often than not, have lower fitness than bearers of a trait variant *T**. According to this definition, *T* could be made dysfunctional by changing the environment in such a way that bearers of *T** proliferate while leaving absolute numbers of bearers of *T* unaffected, thus reducing the relative frequency of *T*. CA_2* disallows such changes in the relative frequency of *T** from changing the functional status of *T*'s activity. Of course, nothing should prevent one from using "functional" and "dysfunctional", in certain contexts, more or less synonymously with "adaptive" and "maladaptive", or simply, "fitness enhancing" and "fitness reducing". But that is not a definition that is appropriate for psychiatry, where "dysfunction" is often used to

characterize a sort of inner state that produces disturbing thoughts, emotions, or behaviors, and not a statistical property of populations.

Given these two conditions of adequacy that will allow for the evaluation of proposed definitions of "function" and "dysfunction", the next section turns to three major attempts on the part of psychiatrists to define "dysfunction" (or related terms).

2.3 THREE PSYCHIATRIC DEFINITIONS OF "DYSFUNCTION"

Of the three definitions of "dysfunction" (or associated terms) that will be presented in this chapter – Kendell's (1975b) "biological disadvantage" criterion, Spitzer and Endicott's (1978) "operational definition", and Klein's (1978) evolutionary definition – the first two are purely consequentialist. In other words, they hold that whether or not a condition is dysfunctional has to do exclusively with the consequences of the condition rather than its causes. The third definition is partly etiological, because it claims that whether or not something is dysfunctional has to do with whether or not it deviates from an evolutionarily determined process. The first two definitions, moreover, do not satisfy CA_2*, because in both, the sort of consequences that the definitions pick out are those that tend to vary depending upon variation in the physical and social environment of the individual, even when this environmental variation has no effect upon the structure or activity of the trait itself. This is not to say that *no* consequentialist definition of "function" is capable of satisfying CA_2*, but it does provides good evidence that consequentialist definitions of "function" will not be suitable to the psychiatric context. Section 3.2 will provide a more careful argument for this claim.

Alternatively, an analysis of Klein's (1978) evolutionary definition will show that definitions based on etiology *are* capable of satisfying both adequacy conditions. Hence this section will provide a motivation for exploring etiological definitions of function in

much more detail. This exploration, along with a more elaborate contrast of etiological and non-etiological definitions of "function", will occupy Chapter 3.

2.3.1 Kendell's (1975b) Definition of Dysfunction as "Inherent Biological Disadvantage"

Kendell (1975b), like Spitzer and Endicott (1978) and Klein (1978), wants to refute antipsychiatric challenges by showing that many of the central disturbances that psychiatrists treat do, in fact, qualify as *bona fide* "illnesses" or "diseases". (Kendell's interest in this challenge is not surprising given his central role in organizing and conducting the US-UK diagnostic study discussed above in Section 2.1.1; see Kendell *et al*. [1971]). His strategy, like that of his American counterparts, is to define "disease" (or "illness") generally and then to show that certain putative mental disorders meet the defining criteria. Although his paper does not attempt a definition of "dysfunction", the definition he offers for "disease" (or "illness") could easily be used to attempt to define "dysfunction", since it is intended to play a similar role in legitimating psychiatry as a medical discipline.

Kendell argues that disease concepts that rely on *etiology* – for example, the idea that diseases are necessarily caused by demonstrable physical lesions – are inadequate because there does not appear to be a uniform etiological pattern associated with all uncontroversial examples of diseases (Kendell [1975b, 308]). Instead, he elaborates a concept introduced by Scadding (1968) that defines disease in terms of its *consequences*, namely, as a statistically abnormal set of organismic characteristics that places its bearer at a "biological disadvantage". Furthermore, Kendell originally defines a "biologically disadvantageous" condition as one that reduces fertility or longevity (Kendell [1975b, 310]) – although he comes to modify this in the course of his argument in order to make it more nuanced.

On the basis of this definition, his argument that schizophrenia is a disease is empirically straightforward. He appeals to post-hospitalization follow-up studies of schizophrenic patients to show that those populations, on average, marry less, and have fewer children, then the general population. Insofar as schizophrenia can be associated with reduced fertility it qualifies as a "disease" in Kendell's sense. He also argues on the basis of intuitive plausibility that homosexuals suffer a reduction in fertility relative to their heterosexual counterparts, and therefore that homosexuality is likewise an illness or disease (Ibid., 311). He also draws upon evidence that people with bipolar disorder (at the time, "manic-depressives") are more likely to commit suicide than their unafflicted counterparts, and that all forms of drug dependence are associated with increased mortality, to argue that both conditions qualify as illnesses (Ibid., 312).

As it stands, Kendell's analysis cannot qualify as an explication of "dysfunction" because it fails CA_2*. By equating a dysfunctional trait with a vulnerability to decreased fertility or longevity, certain conditions can come to qualify (or be disqualified) as dysfunctional owing to changes in features of the environment that have no actual effect upon the characteristic structure or activity of the trait itself, and hence his definition admits of externalist ascription of "disease". For example, schizophrenia may come to be disqualified as an illness because of an absolute decrease in the reproductive rate of non-schizophrenic populations from which the former are isolated.[57]

Kendell does not recognize that his definition admits of externalist ascriptions of disease, although he does recognize a different problem, which is that certain conditions

[57] A recent film, *House of Fools*, DVD, directed by Andrei Konchalovsky (2002; Hollywood, CA: Paramount Home Video, 2004) suggests a rather extreme, but amusing, thought-experiment. It features a mental hospital in the Chechen countryside which comes to be overshadowed by war, and in which, presumably, the average life expectancy of the patients comes to surpass that of the soldiers falling in battle around them. In doing so, the film raises the question of what "madness" consists of; ironically, according to Kendell's definition, the soldiers would qualify as having illnesses, and the hospitalized patients would not.

may come to be associated with decreased fertility or longevity – and hence become "diseases" in his sense – merely because of the stigmatization associated with the mental disorder label, and hence that his definition does not adequately respond to antipsychiatrists' challenges. As Kendell notes, sociologists such as Scheff would argue that, "the main reason people labeled as schizophrenics have relatively few children is because they are regarded, both by others and by themselves, as lunatics and are less likely to marry and have children for this reason…"(Ibid., 313). He also acknowledges that because humans are social animals, it is difficult, if not impossible, to separate those vulnerabilities that are due to negative social attitudes and those that are due to biological causes.

Nonetheless, Kendell maintains that if one wants to figure out whether or not something is truly a disease, one must determine that the relative decrease in longevity or fertility associated with the condition is not merely caused by the stigmatization associated with labeling or by other effects of negative social attitudes about it. Yet, as indicated above, the question of whether the vulnerability associated with, for example, schizophrenia has a biological or social cause is not the real problem in formulating a definition. Rather, the question of definitional adequacy concerns whether the concept of dysfunction can be determined on purely externalist grounds. Once it is accepted that the core problem is externalism about the concept of dysfunction, the whole challenge of attempting to disentangle "merely social" from "exclusively biological" causes of vulnerability is irrelevant to the task. Vulnerability alone, whether due to social or biological causes, should not be considered to be part of the definition of "dysfunction", although it may qualify as a defeasible indicator for the presence of a dysfunction.

Nonetheless, because Kendell thinks that mere social factors should be irrelevant in determining whether or not something is a "disease", he attempts to narrow down his

definition to exclude them. According to his revised definition of "biological disadvantage", a condition places its bearer at a biological disadvantage not merely if it is associated with reduced fertility or longevity; in addition, the cause of this decrease must "be innate and not simply one that leads to rejection by others" (Ibid., 314). Thus the challenge he poses to those who wish to evaluative the disease status of a condition is the following:

> The criterion must be, would this individual still be at a disadvantage if his fellows did not recognize his distinguishing features but treated him as they treat one another? In the case of schizophrenia the argument hinges on whether the high mortality and low fertility associated with this condition are innate, or whether they would melt away if those whom we call schizophrenics were not merely treated like other people but not even recognized as deviant. (Ibid.)

In short, according to Kendell's revised concept of disease, a disease is a condition that places its bearer at an "*intrinsic* biological disadvantage" (Ibid., emphasis added), rather than one due to social causes.

Kendell's view faces one of two major problems, depending on how one chooses to interpret the concept of "innate" or "intrinsic" that he appeals to. One plausible interpretation is that a condition is "innate" if one is born with a predisposition for having it. This interpretation is supported by the relevance to his definition of "disease" that he attributes to evidence for genetic transmission of schizophrenia. This evidence, he claims, "establishes beyond doubt" (Ibid.) that the disadvantage associated with schizophrenia is not merely a consequence of the social stigmatization of people with schizophrenia, and therefore that schizophrenia is "really" a disease. This interpretation, of course, entails that his definition is no longer purely consequentialist but incorporates etiological components as well. But even if one sets aside empirical doubts about evidence for the genetic transmission of schizophrenia, the problem with this interpretation is that the

relevance of this etiological claim to his conclusion has not been established. Even if one supposes that schizophrenia is associated with decreased fitness, and that models for genetic transmission of schizophrenia are valid, in order to know whether or not schizophrenia is a disease one would have to determine the relation between the two. In other words, the problem that Kendell poses cannot be resolved by determining whether or not schizophrenia has a genetic component. Rather, one must determine why it is biologically disadvantageous to have, regardless of its origin. One may be genetically disposed to schizophrenia, yet it may be disadvantageous in some environments for external reasons, e.g., due to the sort of social stigmatization that he believes should not play a role in the determination of disease.

A second plausible interpretation of "innate" or "intrinsic" is suggested by its contrast to "relational". For example, Kendell uses homosexuality as a paradigm case of a condition that is "intrinsically" disadvantageous, presumably for something like the following reason: a mere explanation of what homosexuality is should lead one to infer, given some very general background knowledge concerning the biological basis of reproduction, that a person who exclusively practices homosexuality will not have any children. Similarly, from a description of catatonic type schizophrenia, as well as some very general background knowledge concerning standard requirements for gaining access to food and mates, one can plausibly infer that a person who spends most of his or her life in that condition will be at a biological disadvantage, relative to their normal counterparts.

These examples suggest that the notion of an "intrinsic" biological disadvantage that Kendell appeals to can be interpreted epistemologically, e.g., in the sense that *that* the condition is biologically disadvantageous is derivable from the definition of the condition, in addition to some very basic empirical knowledge about the activities that are

necessary to one's survival and reproduction. But is important to point out that conceptually, this criterion is still relative to the current environment of the person so affected, and hence is capable of being externalistically individuated. There are probably very few mental disorders, if any, which absolutely interfere with survival or reproduction. The disadvantage remains relative to the survival and reproductive rates of individuals within the same population. Hence the distinction between "intrinsic" and "relative" biological disadvantage is not a qualitative one.

This epistemological interpretation of "intrinsic" is also relative to one's current environment in a second way, namely, to the background knowledge that can be legitimately included in its assessment. Because it incorporates a reference to the background knowledge that one has about a condition, then changes in the amount and type of that knowledge can change whether that condition is "inherently" or "not inherently" biologically disadvantageous. For example, someone who only knows that homosexuality involves a disposition to sexual attraction towards members of the same sex cannot derive anything about whether or not homosexuals tend to leave more, less, or the same number of offspring as others who are not so disposed. Consequently, on either interpretation of "intrinsic" or "innate", Kendell's notion of disease as "intrinsic biological disorder" does not satisfy CA_2* in that it does not exclude externalist ascriptions of disease.

Spitzer and Endicott (1978) suggest a third definition of "intrinsic", which was noted above in Section 2.1.2. According to their definition of "mental disorder", which will be discussed in the next subsection, a disorder must be "directly and intrinsically associated with distress, disability, or certain other types of disadvantage" (Ibid., 18). In turn, to say that a condition is "intrinsically associated" with these consequences is to say

that it is associated with them "in all environments" (Ibid., 19)[58]. Furthermore, Spitzer and Endicott mark the distinction between a "disorder" and a (mere) "vulnerability" precisely in terms of whether the condition is disadvantageous in all environments or only in some environments. Yet if a "disadvantageous condition" is interpreted as it is in Kendell (1975b), as one that reduces one's chances of survival or reproduction, then the concept of a condition that is biologically disadvantageous in all environments would seem to be a contradiction in terms, since disadvantage, as noted above, is relative to the survival and reproductive rates of other members of the population. Perhaps the locution could be interpreted as referring to all and only conditions that are absolutely prohibitive of survival or reproduction, such as suicide or infertility. But such an interpretation would render entirely questionable the conceptual or clinical utility of the recommendation, since very few conditions would meet such stringent criteria.

Kendell (1986) has come to reject his prior definition of "disease", and, furthermore, argues that at present no adequate definition exists, though that it would be desirable to find one in order to resolve diagnostic controversies in a rational manner. In criticizing his prior attempt, he offers some fairly straightforward arguments that are worth noting here. He suggests that the core concept of disease attempted there rested upon the concept of "impairment of function", and the notion of functional impairment of a *part* of an organism seems to rest upon whether or not it contributes to the functioning of the organism as a whole (Ibid., 32). But because organisms are not artifacts, it is unclear what their "essential functions" are. In the case of non-human animals, people typically suppose that survival and reproduction are essential functions, and therefore that the appropriate criteria for evaluating such impairment is a reduction in their chances for

[58] The authors qualify this statement to exclude those environments that are specifically created to manage or compensate for the condition.

survival and reproduction (as in Kendell [1975b]). Humans, on the other hand, are inclined to assume that they were "meant" or "designed for" loftier goals, and that:

> [W]e were designed to fulfill a far wider and more exalted range of functions than merely reproducing the DNA of our germ cells. We believe that we were in some sense 'meant' to be honest and trustworthy, to form stable, monogamous relationships, to be rational and even-tempered, intelligent and sensitive, and that any serious lapse from these self-imposed norms is due to a dysfunction or disorder of some kind to which we have attached labels like psychopathy, character neurosis, dyslexia, and pyromania. (Ibid., 35)

Consequently, the concept of "impairment of functioning" does not provide the sort of purely "biomedical" rather than "sociopolitical" criterion of disease that he had wished for (Ibid., 25), since Kendell believes that the concept of human functioning it presupposes is relative to social ideals concerning human well-being or the "good life".

Even supposing, however, that survival and reproduction are sufficiently uncontroversial goals to impute to human beings, he deems his former concept of "intrinsic biological disadvantage" to qualify neither as a necessary nor sufficient condition of "disease". He argues against it by invoking intuitively plausible counterexamples, rather than by appealing to a small set of principled adequacy conditions. It is not sufficient, he claims, because it ignores human volition: a man, for example, may choose not to reproduce because he wants to become a monk, thereby placing himself at a biological disadvantage. Yet for that reason alone he is not ill. Consequently, placing oneself at a biological disadvantage does not suffice for having a disease (Ibid.). (Klein [1978, 64] similarly rejects Kendell's (1975b) definition on the grounds that it would make the rational use of contraception a disease.) On the other hand, he claims, being placed at a biological disadvantage is not necessary for having a disease, since there are several conditions that are uncontroversially regarded by

physicians as diseases, such as psoriasis and chicken pox, which do not necessarily place their bearers at a biological disadvantage.

However, it should be borne in mind that the ability to discover a small set of isolated counterexamples to one's definition should not suffice to invalidate the definition or to exclude it from any further consideration. This is particularly true in the case of psychiatry, when there are very few uncontroversial exemplars of mental disorders that one can reliably invoke. Hence, the analytical strategy followed by this dissertation will typically involve, as has been the case in this section, the defense of a small set of principled adequacy conditions and the evaluation of proposals on that basis rather than the reliance upon the more *ad hoc* approach exemplified in Kendell (1986).

2.3.2 Spitzer and Endicott's (1978) Operational Criteria for "Dysfunction"

For Spitzer and Endicott (1978), the concept of disorder (whether mental or nonmental) includes three main components: an inferred internal dysfunction, negative consequences of the condition, and a "call to action". This call to action is addressed to society, to provide medical assistance for people with the condition, and to the individual so afflicted, to adopt the "sick role". These three components, however, are not independent of one another in their view. For if the "call to action" is truly warranted, the condition must be seen as having negative consequences that stem from an inner dysfunction: "Implicit in the call to action is the assumption that something has gone wrong within the human organism which has led to negative consequences" (Ibid., 18). Hence, similar to the view expressed in Klein's (1978) analysis of "mental illness" (see Section 2.3.3), the presence of an inner dysfunction that produces negative consequences is crucial for legitimizing the "sick role", presumably because it carries the connotation that the person in whom the dysfunction resides is not responsible for those negative consequences, but rather, is "afflicted" by them – the "patient", rather than the agent; or,

the one who suffers, rather than the one who acts. Consequently, due to its prescriptive connotations, the authors recognize the importance of providing some explanation of the conditions under which something is considered a dysfunction.

However, although Spitzer and Endicott provide an explicit definition of "disorder" and of "mental disorder" (as cited above in Section 2.1.2), they do not attempt an explicit definition of "dysfunction". Rather, they claim to provide "detailed operational criteria" (Spitzer and Endicott [1978, 17]) for applying their definition of disorder, and the satisfaction of these criteria is supposed to imply there is such a dysfunction. An advantage that they claim to be associated with providing operational criteria for their definition is that it allows them to avoid the perplexing and controversial task of providing explicit definitions of terms such as "dysfunction", "maladaptive", and "abnormal", which often figure into other definitions of "mental disorder" (Ibid., 17). In order to evaluate the adequacy of their attempt, then, some elaboration of the notion of an "operational definition" is warranted, since, if this can be successfully carried out, then the whole project of constructing an explicit definition (one replete with necessary and sufficient conditions) for "dysfunction" can be avoided.

In its classical sense, an "operational definition" for a term is a definition that consists of a set of procedures, or "operations", that any sufficiently informed and competent person can perform to determine the applicability of the term (e.g., Bridgman [1936]). A conventional example is "length". On the one hand, one might try to define "length" explicitly as, e.g., "a measure of spatial extension", although, if one were asked to define "spatial extension", one would rapidly end up in very perplexing and philosophically controversial territory. Is spatiality a phenomenal concept, a physical concept, or something different altogether? An operational definition of the term, on the other hand, consists of a set of instructions for determining the length of an object, e.g.,

"lay a rigid rod of type *R* end to end along the object in question and count how many times you perform this procedure. If you have done this *n* times, then the length of the object is *n R*'s." So long as one understands how to carry out the procedure, any further conceptual analysis of "length" is unnecessary.

Furthermore, an operational definition of a term is intended, in the strictest sense, to qualify as a *definition* of the term, rather than, say, a defeasible indicator for the applicability of the term. For example, if one were given an operational definition of determining the truth of any statement of the form, "Object *x* is *n* meters long", and one were to reply by asking, "Yes, but what does it *really* mean to say that the object possesses length, over and above the fact that such an operation yields certain results?", one's response would be confused or meaningless. The specified procedure completely and exhaustively determines the applicability of the term.[59]

Although the foregoing explanation may seem to belabor the point, it is important to clarify precisely the status of the criteria that Spitzer and Endicott offer. For those criteria, by their own admission, do not qualify as an operational definition of "dysfunction", but rather, a set of indicators that psychiatrists often take to strongly suggest the presence of an organismic dysfunction. Moreover, they are not really "procedures" in the narrow sense in which the term is typically used, that is, to stand for a series of physical or symbolic manipulations. For although some of their indicators for whether or not a person has an inner dysfunction qualify as "procedures" in this sense, others demand a more introspective approach on the part of the psychiatrist regarding the extent to which he or she is capable of sympathizing with the goals of the person whose

[59] One of the main problems with operationalism is precisely the fact that what appears to be a single concept, such as weight, may be associated with several different types of quantitative measurements, which suggests a difference between the concept itself and its associated measures. The operationalist, in turn, typically rejects this appearance of unity, and insists that, e.g., each usage of the expression "weight" in "atomic weight", "weight of a person", and "weight of a planet", stands for a different concept.

psychiatric condition is under evaluation. Nonetheless, even granting this broad sense of "procedure", two consequences immediately follow concerning their proposal.

The first is that, since they do not offer a definition of "dysfunction", operational or not, the task of explicating "dysfunction" remains an open one. This is not to say that, for all practical purposes, the criteria they propose are insufficient for many of the diagnostic and classificatory purposes to which they may be put. It is also not to say that a "good science" requires rigorous definitions, operational or not, for all of the terms in its specialized nomenclature. Nonetheless, in the absence of such an explication, one cannot evaluate the extent to which the use of "dysfunction" in the psychiatric context is sufficiently constrained to permit a clear distinction to be drawn between mental disorders *proper* and any significantly socially-disvalued psychological or behavioral condition. But drawing this distinction constitutes the motive for explicating the term in the first place!

The second is that, since they rely upon the sympathetic evaluation of the consulting clinician in determining whether or not something is a mental disorder, the criteria admit of externalist ascription and hence do not satisfy CA_2*, since the extent to which a psychiatrist is capable of sympathizing with another can change in ways that are independent of changes within the other person. This provides a further reason why these criteria should not be thought to constitute part of the definition of the term "dysfunction". However, as will be described below, the elaboration of the criteria they offer may still be useful in imposing reasonable constraints upon a definition of "dysfunction", and hence they will be elaborated in some detail.

After providing an explicit definition of "disorder" (quoted in Section 2.1.2) they propose four criteria to determine its applicability more precisely and concretely. The first criterion (A) specifies that the condition in its developed form must be associated

with distress, disability, or disadvantage, in all environments (besides the one specifically created to compensate for the condition) (Ibid., 19). This criterion guarantees that the condition is associated with the sorts of negative consequences that warrant the sick role. They also provide a list of six specific disadvantages (a through f) any of which, they claim, is presently considered, "in our culture, as suggestive of some type of organismic dysfunction":

> a. Impaired ability to make important environmental discriminations
> b. Lack of ability to reproduce
> c. Cosmetically unattractive because of a deviation in kind, rather than degree, from physical structure
> d. Atypical and inflexible sexual or other impulse-driven behavior which often has painful consequences
> e. Impairment in the ability to experience sexual pleasure in an interpersonal context
> f. Marked impairment in the ability to form relatively lasting and nonconflictual interpersonal relationships (Ibid., 20).

The reason that the authors explicitly relativize this list of "disadvantages" to a specific culture at a specific time is because they recognize that the types of conditions that are recognized as disadvantageous are susceptible to change depending upon the importance that a given culture places upon the associated normal behavior. Therefore, according to Hare's distinctions raised in Chapter 1.3, "disadvantage" as they use it is an evaluative term. But this implies that its application is subject to change as a function of changing values and consequently its ascription is externalist, and hence it does not satisfy CA_2*. For example, Spitzer recognizes that claiming "impairment in the ability to experience sexual pleasure in an interpersonal context" as a significant disadvantage is, in part, a politically motivated effort to exclude homosexuality as a mental disorder but include fetishism and other forms of sexual arousal that do not involve other human beings as mental disorders, and that whether one considers this inability to qualify as a

disadvantage is relative to the value one places on the associated ability (Spitzer [1981, 212]; also see Section 2.1.2 on Spitzer's introduction of the concept of "disadvantage"). Hence criterion (A) does not belong within a definition of "dysfunction", operational or not.

The second criterion (B) is more interesting in this context because it represents, along with the third criterion (C), their attempt to provide more specific indicators for the presence of an organismic dysfunction, or to justify the expression that "something is wrong within the organism" (Ibid., 26). According to criterion (B), "the controlling variables [of the disorder] tend to be attributed to being largely within the organism with regard to either initiating or maintaining the condition" (Ibid.). The intuition that this criterion is supposed to capture is that nothing has necessarily "gone wrong within" the organism if that organism's disturbing or unusual behavior is merely a response to some transitory environmental contingency, or that it is merely due to "noxious environmental influences" (Ibid., 28). Puzzlingly, they say that, "the violation of this principle results in labeling dissidents in certain countries as mentally ill on the basis of their inability to conform to the political and social norms of a particular repressive society" (Ibid.). Presumably, their idea is that if one revolts against a repressive society, then one's actions are spawned more by noxious environmental stimuli rather than some inner mechanism that disposes one toward exhibiting rebellious tendencies and that it should therefore not qualify as stemming from an inner dysfunction.

The third criterion, (C), is that a condition is not included if the negative consequences of the condition are "apparently the necessary price associated with attaining some positive goal" (Ibid.). This provision excludes childbirth from qualifying as a disorder, as well as the grief following the loss of a loved one, insofar as the pain of childbirth is the price one pays for children, and grief is the price one pays for having

attachments, and the goals of having children, or emotional attachments to others, are positive ones. This criterion captures the intuition that, "when individuals undergo deprivation and distress in order to obtain some understandable positive goal, we assume that the organism is working and do not infer a dysfunction" (Ibid., 29). However, this raises the fairly obvious objection that the distress associated with the pursuit of *certain* goals by a person can, in and of itself, indicate a mental disorder, such as the goal of avoiding former friends and loved ones if it is motivated by the unwarranted belief that they wish to harm one. Hence the authors point out that, "the distress is less likely to be considered as due to a mental disorder to the extent that the positive goal is understandable and in keeping with reality" (Ibid., 29). But the fact that this indicator for the presence of an internal dysfunction relies upon a person's capacity – not necessarily the one that has the alleged dysfunction – to sympathize with the goals and motivations of another implies that the criterion admits of externalist ascription, and hence, like criterion (A), if it were to be adopted as a component of an operational definition for "dysfunction" it would not satisfy CA_2*, in the sense that its applicability can change by changing the psychological state of the evaluating psychiatrist, where this change need not have any effect on the person whose condition is being evaluated.

The fourth criterion, (D), states that the condition must be distinct from others in at least one of the following variables: "clinical phenomenology, course, response to treatment, familial incidence, or etiology" (Ibid.). The purpose of this condition is to ensure the applicability of the so-called "medical model" as defined by Spitzer *et al*. (1977), which is the hypothesis that there are "organismic dysfunctions which are relatively distinct with regard to clinical features, etiology, and course" (Ibid., 5).

The main consequence that follows from these considerations about the authors' proposed criteria is that, as noted above, because the criteria consist of defeasible

indicators for the presence of an inner dysfunction, it does not qualify as a definition of the term, and hence does not satisfy the motivation for explicating the term. Nonetheless, the criteria they propose are useful because they express, as a matter of descriptive sociology, the sorts of considerations that psychiatrists have implicitly used, and, perhaps, continue to use in practice, to infer the presence of a dysfunction. Hence, these criteria may be useful for placing some reasonable constraints on any definition of "dysfunction" that is suitable for the psychiatric context, namely, that to the extent that psychiatrists *do* implicitly apply these criteria, it ought to make sense of why they do so, that is, why they tend to take these criteria as valid indicators for an inner dysfunction. That does not mean that a definition of "dysfunction" should assume that these indicators really are valid, in the sense that they actually measure the construct their use is intended to measure, but that it should make sense out of why they are often thought to be valid. For example, by definition, if something within a person is "malfunctioning", then it is not just the case that it is *not* currently performing its function, but that it is in some sense *incapable* of doing so, even in the range of environments within which, historically, it came to possess that function. As a consequence, one would expect Spitzer and Endicott's criterion (B.1) to be a reliable indicator of an inner dysfunction (along with the other criteria), namely, that "simple informative or standard educational procedures do not lead to a reversal of the condition" (Spitzer and Endicott [1978, 27]). This is because the incapacity of a trait to perform its function is not typically a consequence of ignorance or mistaken beliefs on the part of the person who possesses the trait, and this is often true, in turn, because the historical entrenchment of functional traits should guarantee some measure of functional autonomy from the conscious volition of the person who possesses them. Wakefield (1993) in an illuminating critique of the same article – despite his gratuitous appeals to organismic "design" – makes a similar point

about their criterion (B.2), according to which, "nontechnical interventions do not bring about a quick reversal of the condition". As he points out, as a matter of empirical contingency, organisms and artifacts "generally function the way they are designed to function" (Ibid., 169). As a consequence, "breakdowns in designed functioning are likely to be less frequent and their solution is likely to involve specialized knowledge, so the appropriate interventions are more likely to be left to a technical elite" (Ibid.). Hence the link between being dysfunctional, and requiring technical intervention, though contingent, ought to be a reliable one. Moreover, insofar as these criteria are not definitional, their practical and clinical utility is not threatened because some of the conditions for their satisfaction are externalist (do not satisfy CA_2*).

2.3.3 Klein's (1978) Evolutionary Definition of "Dysfunction"

One of the recurring themes within the foregoing definitions of "dysfunction" is that having a dysfunctional condition (disease, disorder, etc.) depends upon the consequences that the condition produces, rather than its history or causes. One of the reasons for these attempts to construct consequentialist definitions of "dysfunction" is that in most cases, the etiology of the mental disorder is unknown, and hence if the definition had to rely on etiology, the warrant for applying the term would be questionable. Even where etiology is known, however – as in the case of substance-induced delirium – or when plausible theories exist to account for the origin of the disorder, the diversity of known or plausible causes does not seem to fit any unique or well-defined pattern that could be encompassed by a single definition. (This was the rationale behind Kendell's [1975b] restriction to consequences.) For some disorders, such as bipolar disorder – at least if one judges by the relative efficacy of pharmacological intervention in ameliorating them – some biological etiology appears plausible; for others, such as posttraumatic stress disorder (PTSD) or adjustment disorder, the etiology

is almost, as a matter of definition, "psychological" in that it refers to the manner in which an agent interprets, or gives meaning to, his or her experiences.[60] This multiplicity of etiological patterns also holds true for nonmental medical disorders as well, some of which stem from a determinable lesion, some from the presence or absence of a specific protein, and others from a quantitative deficit of a vital nutrient.

Nonetheless, consequentialist definitions of "dysfunction" are also problematic, since the sorts of consequences that a condition gives rise to, and, more importantly, how those consequences are interpreted and valued, are often so contingent upon the specific environment within which the afflicted person finds himself or herself that, in the final analysis, whether or not the condition qualifies as dysfunctional (disordered, etc.) need not be determined by, or supervene upon, changes that take place within the individual himself or herself.

The approach taken by Klein (1978) differs from the foregoing attempts in that it is explicitly etiological in character (although it does contain some consequentialist components). Reading him generously, he defines the "function" of a trait in terms of its evolutionary history, and specifically, in terms of its selection history. Then he defines "dysfunction" as a suboptimal deviation from this function. Hence, for a trait to be dysfunctional it must be unable to perform the activity for which it is an adaptation.

Two features of this definition should be remarked upon. The first is that the etiological approach exhibited by Klein's definition reveals an interesting conceptual relation between two apparently disparate facets of the prior, consequentialist, attempts to define "dysfunction". The first facet is that although they both proved inadequate because they failed to satisfy CA_2*, nothing was stated concerning whether or not they satisfy

[60] This does not, of course, mean that psychological and biological models are mutually exclusive or that disorders do not emerge from the interaction of several types of processes; see Bolton (2003) for discussion and references on the interaction between biological and psychological causes of PTSD.

CA_1. The reason for this is that neither Kendell (1975b) nor Spitzer and Endicott (1978) attempt to define the corresponding concept of "function". They both, as it were, skip ahead to define its negative counterpart, "dysfunction", directly, rather than to build upon the intuition that being dysfunctional involves a privation or aberration of normal functioning, and that its definition should reflect this dependence relation. The second feature of these attempts is that both are motivated by despair of finding *any* unique etiological pattern which can encompass all of the diverse ways in which something within the organism can "go wrong", and hence conclude that etiological definitions will be hopelessly inadequate to the task. Yet if the problem of defining "dysfunction" is transformed into the problem of defining "function", and if the former is defined in terms of the latter, there need not be any unique etiological account of the way in which a functional trait becomes "dysfunctional". An organ, for example, can "malfunction" because of a lesion, a microbial infection, or by a deficit of some vital nutrient – so long as any of these causes impede its normal function from being carried out, it can rightly be called "dysfunctional".

The second major feature of the etiological approach exhibited by Klein's definition is that it is, in general, capable of satisfying both adequacy conditions (despite the fact that one of the elements of Klein's particular formulation appears to fall afoul of CA_2*, as will be discussed below). On the one hand, it is capable of satisfying CA_1. As noted above, in order for a definition of "function" to lend itself to a corresponding definition of "dysfunction", it must allow a conceptual distinction to be drawn between *having* a function and *performing* a function, since being dysfunctional involves having a function that one cannot perform. If the function of a trait is determined by its history – for example, that for which it is an adaptation – then whether or not a trait has a function is conceptually independent of whether it is currently capable of performing that

function. Another way of putting the point is that having a function supervenes upon one's history, and performing a function upon one's current structure and activity.

On the other hand – though this is more problematic – it is capable of satisfying CA_2*, since whether or not a trait is dysfunctional depends upon whether or not it is impeded in performing the activity for which it is an adaptation, and this often, although not always, supervenes upon "inner" changes – namely, either those involving the material of which the trait is composed (assuming it to be a physical trait), the way in which that material is structured, or the way in which that structure physically interacts with other structures within the same organism. For example, if the function of the heart is to beat, then whether or not it is dysfunctional is determined, at least in part, by whether or not it is beating, and this capacity trivially supervenes upon the heart's activity and structure rather than external factors such as changes in social attitudes about hearts.

It is arguable that, in some cases, according to this definition, a trait becomes dysfunctional exclusively as a consequence of "external" factors and hence does not satisfy CA_2*. For example, it may be that the function of sperm is to fertilize ova, and that a given sperm is impeded in its capacity to do so because of a woman's use of oral contraception, in which case the sperm cannot perform its function even though no structural change distinguishes it from a functional sperm. Nonetheless, it is not as conceptually difficult to restrict the etiological definition of "function" in such a way as to exclude such counterexamples, as it is to restrict the consequentialist definitions in a similar manner. In fact, the purpose of Chapter 3 is to argue that of the major explications of "function" that have been proposed by philosophers, *only* etiological accounts are capable of satisfying both adequacy conditions. This claim is not, of course, equivalent to the claim that it is not *possible* for a non-etiological account of function to satisfy both

adequacy conditions or that non-etiological accounts do not have the "conceptual resources" to do so, but it will provide compelling evidence for the latter.

What follows is an overview of Klein (1978), an elucidation of the main problems it confronts, and potential solutions to those problems. For Klein, like Spitzer and Endicott (1978), the notion of a mental disorder incorporates two different components that stand in a certain relation to one another. The first component is a sociological one: a person with a mental disorder occupies the "sick role". The sick role can be self-ascribed or assigned by others. The second is that there is an involuntary impairment in organismic functioning (Klein [1978, 70]). Again, like the DSM-III characterization, this impairment is thought of as a property of the individual that is not relative to the individual's current social environment. The relation between the two, however, is not a *causal* one; rather it is a legal or ethical one: the involuntary impairment (or organismic dysfunction) *legitimately entitles* a person to occupation of the sick role.

The position of the "sick role" within the hierarchy of social roles will be briefly outlined. A "socially-defined role" implies a "system of rights and duties" that "defines the expectations regarding the action of others toward you and vice versa" (Ibid., 42). Role relationships can be "exploitative" or "fair", the latter implying reciprocal exchange of benefits. One type of exploitative social role is the "exempt role", in which the person exempts himself or herself (or is exempted by others) from normal obligations and duties. The occupation of the exempt role can be legitimate or parasitical. The "sick role", then, is a type of legitimate exempt role, and the legitimacy of occupying the sick role is bestowed by involuntary impairment.

What is interesting about Klein's definition is that it approaches the notion of a mental disorder from the standpoint of a problematic claim to exemption from reciprocal rights and duties. In other words, Klein conceives of the clinical decision procedure in

psychiatry *not* as the attempt to differentiate between mental disorder and mere "social deviance", but the attempt to differentiate legitimate and parasitical claims to the exempt role. Both of these roles entail exploitative role relationships. The import of this perspective is that regardless of whether the person in question truly has a mental disorder, or merely exhibits failure of social-role functioning, the behavior that eventuates in clinical psychiatric consultation is exploitative and therefore ought to be changed.

Klein observes that it is problematic to define "disease", "illness", or "disorder" merely in terms of its undesirable clinical manifestations, since what are considered to constitute "undesirable manifestations" tend to vary with the values and ideals of a given social group. At the extreme, the person who accepts this analysis of "illness" would have to accept Sedgwick's (1973) conclusion that, "there is no mental or physical illness in nature…The human evaluation of certain conditions and being deviant and undesirable leads to their segregation as states of illness" (cited in Klein [1978, 45]). The problem with Sedgwick's conclusion, however, is that it misses the "necessary crucial inference" behind the attribution of illness, namely that "something has gone wrong [within the individual], not simply that something is undesirable" (Ibid., 46). Thus, the problem that Klein confronts is that of defining a concept of disease that is conceptually independent of all and any of its negatively-valued consequences:

> [C]an one define disease in a fashion conceptually independent of illness? Is there a positive, scientifically definable criterion for pathogenic process or disease? Can one distinguish an abnormal disease state from simple biological variation without using illness as a necessary condition? Unless we can do this we will not be able to meet Sedgwick's criticism that all illness and disease categories fundamentally represent nothing but arbitrary social evaluation. How do we know that something has gone wrong? (Ibid., 49)

Such a concept of disease, Klein thinks, is available, and can be drawn from "the systematic implications of modern biology" (Ibid., 45). Specifically, the inference that "something has gone wrong" within the person can be justified through evolutionary reasoning. Insofar as organismic forms are adapted, through natural selection, to perform certain functions that contribute to their survival and reproductive capacity, then a *dysfunctional* condition can be equated with the incapacity of a trait to perform the activity that it was selected for. More precisely, a dysfunctional state is "a suboptimal deviation from [an] evolutionary determined process" (Ibid., 51).

There are two problems that arise with his formulation of his position. The first is that, although he states that evolutionary theory is relevant to the determination of function, he never provides an explicit definition of "function", and hence his usage allows for multiple cogent interpretations. For example, he writes that, "species develop a variety of ancillary biological equipment and practices, through variation and selection, fulfilling specific adaptive functions" (Ibid., 50). These adaptive functions include, for example, energy gathering, information gathering, and others which, typically, "serve an overall organismal goal of survival" (Ibid., 51). One plausible interpretation of his usage of "function" that this passage suggests is that the function of a trait should be defined in terms of its evolutionary history. According to this etiological definition, a trait currently has the function of performing an activity only if the activity of that trait has, historically, been selected for by natural selection over other variants of that trait. Consequently, such function ascriptions involve commitments to hypotheses concerning actual historical circumstances.

Yet his comment is also consistent with a different definition of "function", according to which the function of a trait merely consists in its capacity to satisfy a specific biological need, such as energy gathering, that can be determined independently

of any knowledge about the actual course of evolution. Certainly, traits that contribute to such vital activities more efficiently or effectively than others will tend to get selected for, and they may be adaptations themselves, yet this would be a contingent rather than a necessary consequence of the fact that it serves an adaptive "function".

This second interpretation of "function" is suggested by the importance that Klein places upon adopting an "engineering perspective" for the determination of function, and specifically, for the determination of a trait's "optimal functioning": "With growing biological knowledge, we will be able to make more and more exact statements concerning optimum part and integrative functioning from an engineering point of view" (Ibid., 52). This suggests that, conceptually, a trait performs an adaptive function insofar as, from an engineering point of view, it exemplifies "good design" for satisfying a given biological need, rather than because it has been selected for that activity.

The reason that the distinction between these two different concepts of "function" is important to emphasize in this context is that it critically determines the extent to which the definition of function satisfies CA_1, and, by extension, CA_2*, in a fairly non-problematic manner. For, as noted above, the evolutionary definition of "function" clearly and non-problematically satisfies CA_1, since the criteria for having a function and performing that function are different: the first relies on evolutionary history; the second on current structure and dynamics. It is unclear how this distinction would be made given the second definition of "function" – according to which something must exhibit "good design" for a specified task – since the latter suggests that something has a function only insofar as it regularly or typically *does* contribute to satisfying some biological need. This, in turn, suggests that the criterion for functioning well *versus* being dysfunctional would require a statistical component, e.g., a token of a given biological type would be dysfunctional only if it is not performing the activity that the vast majority of tokens of

that type perform in a manner that exemplifies good design from an engineering point of view. Yet this criterion would fail CA_2*, since, as noted above, statistical abnormality is an externalist criterion for marking the distinction between a dysfunctional and non-dysfunctional token of a trait.

Despite the ambiguity in its formulation, then, it will be assumed that Klein proposes an etiological definition of function. This is also consistent with a much later article by Klein (Klein [1999, 423]), in which he largely endorses Wakefield's (1992a; 1992b) "harmful dysfunction" analysis of disorder (see Section 1.3) according to which a "disorder" is a "harmful dysfunction", and "dysfunction" is analyzed etiologically, as a deviation from an evolutionarily determined function. Insofar as it is etiological, Klein's definition of function satisfies CA_1.

However, his formulation, as it currently stands, does not clearly satisfy CA_2*. This is because he defines "dysfunction" as a *suboptimal deviation* from an evolutionarily determined process, and hence its application presupposes an assessment of "optimal functioning". Yet according to Klein, standards of optimal functioning are relative to the specific environment of the organism: "Optimum functioning can be defined only with regard to specified environments" (Klein [1978, 52]). For example, he writes, "it seems evident that certain physiological functions are optimum and adaptive to the Arctic but would be suboptimum and maladaptive for the tropics, e.g., spheriodal versus elongated body" (Ibid.). But in the absence of a specific criterion by which "optimal" functioning can be determined, one cannot evaluate whether or not this standard satisfies CA_2*. For example, if the optimal level of functioning for a given trait of a given individual is determined by the range of variation that is actually exhibited across the population in question, then the assessment of optimal functioning admits of externalist elements and hence does not satisfy CA_2* – since whether or not something within the individual is

functioning "optimally" depends upon what the other individuals are doing. But if the range of variation in performance according to which "optimal" can be identified is not determined by the range of performance actually exhibited by the members of a given population, then the basis for this standard of optimal performance must be specified.

Although Klein recognizes that this relativization to the environment "makes for great complexity" (Ibid.), he does not think that it essentially undermines the definition since merely relativizing standards of "optimal functioning" to the current environment is different from reducing function ascriptions "to the arbitrary category of pure idiosyncratic evaluation" (Ibid.). Certainly, Klein is correct that being environmentally-relative is not the same as being subjective. Yet unless the environmental changes have an effect upon the characteristic structure or activity of the trait in question, then the claim that a given trait's activity is dysfunctional admits of externalist determination, and that is what his definition must avoid.

Of course, this inability to satisfy CA_2* is not an insurmountable problem for etiological dysfunction ascriptions *as such*, but only for Klein's proposal. For example, one can define a "dysfunctional" trait as one that is incapable of performing its function *even in those environments within which it came to possess that function*. (This is precisely the strategy that will be adopted in the next chapter.) This modification prevents a functional trait from becoming "dysfunctional" merely because the bearer of the trait in question has been placed into an environment to which it is not habituated.[61] This definition of "dysfunction" has the advantage that it can distinguish a dysfunctional trait from one that is incapable of functioning merely because it is within a highly abnormal environment. Moreover, this supports the intuition expressed in Spitzer and Endicott

[61] That etiological function ascriptions should be relativized to the "normal" environment of the trait, where this norm is determined by the range of environments within which the trait came to possess the function, is made explicit in Millikan (1984, 33).

(1978) that one should distinguish between the case in which something is really "going wrong within" the organism and the case in which the abnormal trait activity is a response to "noxious environmental stimuli". Finally, this restriction is consistent with Klein's own qualification that a trait should not be considered "dysfunctional" if its suboptimal performance according to one criterion is a result of the fact that it is actively compensating for impairment of a different trait (Klein [1978, 52]) since this suboptimal performance on one criterion may be a highly adaptive response to an abnormal environment. In other words, supposing a trait to be unable currently to perform its function, the judgement that the trait is "dysfunctional" involves, in addition, a negative answer to the counterfactual question: given the current structure and activity of the trait, were it to be situated in its normal environment, would its still be incapable of performing its function? Note that the concept of "environment" here refers not only to the external environment of the organism that bears the trait but also to the internal physiology of the organism within which the functional item is embedded.

Many of these modifications to the etiological definition of function that Klein proposes will be elaborated and defended in detail in the following chapters. The purpose of its introduction here is to suggest the rich conceptual resources that etiological theories of function permit in satisfying the two adequacy conditions that have been imposed upon any definition of "function" that is relevant to the psychiatric context. In the next chapter, etiological and non-etiological definitions of "function" (and "dysfunction") will be elaborated and a more rigorous analysis of the extent to which etiological definitions of function uniquely satisfy both adequacy conditions will be provided.

Chapter 3: From Internal Dysfunctions to Etiological Functions

The purpose of this chapter is to argue that the etiological theory of function is uniquely capable of satisfying both adequacy conditions outlined in the previous chapter (Section 2.2.2). The first adequacy condition (CA_1) is that any definition of function must lend itself to constructing a corresponding definition of "dysfunction"; the second (CA_2*) is that it must do so in such a way that the difference between a dysfunctional and non-dysfunctional entity is not determined by externalist criteria. According to the etiological theory of function, to ascribe a function to an entity is to say something about its history; specifically, it is to provide an explanation for how that type of entity came to exist, or why it continues to exist. The next chapter will defend the appropriateness of a specific version of the etiological theory of functions, namely, the "strong persistence-based etiological theory", according to which the function of an item is that effect that it was selected for by some selection process and which thereby explains the persistence or reproduction of that type of entity within a population of such entities.

The first section (Section 3.1) will provide an overview and taxonomy of current philosophical theories of function (that is, explications of what "function" means). At the most general level, there are two main theories of function: those according to which the function of an entity is determined exclusively by its causal history (etiological) and those according to which it is determined, in part, by the consequences that the entity produces (consequentialist). Within each of these categories there are several subdivisions that depend upon the precise way in which the term is defined.

The taxonomy that will be presented will not be *complete* in the sense of categorizing all *possible* theories of function. For example, one could invent a theory of function according to which the function of an entity is determined by history, yet the

past events that determine the function have no causal relationship with the currently existing entity. The taxonomy that will be presented will have no category for such a theory. This is because the only reason that theories of function that refer to past events are contrived is because they are intended to be explanatory, and so it is not clear what would motivate a theory of function that refers to past events that do not explain the present existence of the entity.

The second section (Section 3.2) will argue that the etiological theory of function is uniquely capable of satisfying both adequacy conditions. It will do so by showing that the consequentialist theories outlined in Section 3.1 violate one or the other conditions, and that etiological theories are consistent with both. It will not provide a deductive argument that it is not *possible* for a consequentialist theory of function to satisfy both adequacy conditions, but rather, it will argue for this claim by showing that well-developed consequentialist attempts to define "dysfunction" fall afoul of one or the other adequacy conditions. In some sense, the arguments presented in this chapter merely represent a systematization of those presented in the previous chapter (Section 2.3).

Are etiological definitions of "function" in general more appropriate for scientific activity than other types of definitions? As noted in the first chapter – and as will be noted below – different uses of the notion of function seem to exist even within biology. The viewpoint adopted in this dissertation, then, is a broadly pluralist one: different concepts of function may be appropriate for different contexts. Although this chapter argues that etiological definitions of function are necessary and sufficient for satisfying the two adequacy conditions specified in the previous chapter, it does not discuss here whether fulfilling these two adequacy conditions is in any ultimate sense desirable or good. This issue will be raised again in the concluding chapter.

3.1 TAXONOMY OF THEORIES OF FUNCTION

As noted above, there are two main approaches to defining "function": those that assume that function ascriptions explain how that type of entity came to exist (etiological), and those that assume that function ascriptions refer only to the consequences that the entity produces (consequentialist). Etiological theories are often called "backwards-looking" theories and consequentialist, "forward-looking" theories. However, since a theory can be both "backwards" and "forward" looking, the following taxonomy will label as "etiological" all theories that contain *only* etiological criteria – that is, those that refer exclusively to history – and it will label "consequentialist" all theories that contain *any* consequentialist criteria (even if they contain some etiological criteria as well).[62] In the first subsection (3.1.1), the problem context that historically motivated the analysis of function statements will be presented, and it will be shown how this context motivates the two main types of analyses that are present in the literature. The second subsection (3.1.2) will classify etiological theories; the third subsection (3.1.3) will classify consequentialist theories.

To render the terminology uniform, the following terms will be used throughout: an *activity* may be the function of an *entity*. Although "entity" usually connotes a spatio-temporally discrete object, the term will be used broadly to refer to activities or qualities as well. This breadth is important because it avoids placing any prior restrictions upon permissible subjects of function ascriptions. Hence one may say that the *heart* (a spatio-temporally discrete individual) has the function of *beating* (an activity), or that the

[62] This convention is the opposite of that utilized in a recent survey of functions (Garson [forthcoming]), in which "etiological" refers to all theories that contain *any* etiological component, and "consequentialist", *only* consequentialist components. This is because, in that context, the focus was on *explanation*; as long as a theory has *any* etiological component it is explanatory. In this context, the emphasis is on *normativity*, and, as will be argued in Section 3.2.1, if a theory of function contains *any* consequentialist criteria then it cannot be normative in the required sense – that is, it cannot define "dysfunction" in a way that satisfies both conditions.

heart's beating (an activity) has the function of *circulating blood* (another activity), or that *dark coloration* (a quality) has the function of *predator avoidance* (an activity). This entity (whether a thing, an activity, or a quality) is usually a *part* of a *system*, but it may not be. Occasionally, in the biological context, *trait* will be used synonymously with *part*, and the individual that possesses the trait the *bearer of the trait* or the *trait-bearer*.

3.1.1 The Problem-Context for Explications of Function

A simple example of a function statement can serve as an introduction to the main problem that such statements confront. Suppose one asks the following question: "Why do polar bears have dense, water-repellent fur?" A common answer might be, "Because fur of that sort helps the polar bear to retain heat". Intuitively, this appears to be a plausible explanation for the fact that polar bears have dense, water-repellent fur. This sort of explanation is often glossed by the statement that, "The function of dense fur in polar bears is to retain heat". Yet it is problematic because the explanation for dense fur refers to an event (heat retention) that dense fur is responsible for bringing about. Causal explanations, at least since the advent of modern science, are constrained by the principle that a temporally prior event explains a temporally posterior event, and not vice versa. Hence it appears that functional explanations cannot be causal explanations for the existence or form of a trait (the problem of "backwards causation"). This gives rise to the question of what, if anything, function ascriptions explain.

Explanations that account for the existence or form of an entity by referring to one of the consequences it produces are called teleological explanations, from the Greek word *telos*, meaning "goal" or "end". Hence function ascriptions are often thought to be a type of teleological explanation; if this is true, then what has been said, historically, of teleological explanations can equally well be said of function ascriptions. Given the problem of backwards causation, any plausible account of "function" must *either* explain

how it can be that the effect produced by a kind of entity can have causal relevance to the existence of the entity, *or* dissolve the misleading appearance that function ascriptions are teleological explanations at all. *Etiological* approaches to function adopt the former route; *consequentialist* approaches the latter.

Intuitively, one might motivate either of the two main approaches to function by considering the following question: what distinguishes a *function* of an entity from a mere *effect* that it produces? To take a trite, but useful, example, why is it said that the (or *a*) function of the heart is to pump blood, rather than to make throbbing sounds? Two different answers present themselves as initially plausible:

(i) according to the etiological view, what distinguishes the *function* of an entity from a mere *effect* is that the capacity of the entity to perform that function explains "why it is there" in that system. For example, it is the capacity of windshield wipers to remove water from windshields that explains why they are on car windshields; i.e., why manufacturers place them there. Similarly, it is the fact that hearts have been selected for because they pumped blood that explains why, presently, creatures with hearts exist. Therefore, in conformity with the logic of teleological explanation, it is true to say that the heart's capacity to pump blood explains why hearts currently exist. However, supposing that the heart was not selected for because of the beating sounds that it makes, there is no sense in which the heart "is there" because of its capacity to make such sounds;[63]

[63] There are, of course, exceptional cases in which it can be said that the heart's beating sounds explain why it is there. For example, if the beating sounds made by a person's heart alert a doctor to a potential heart problem that is thereby remedied, then one can say that the heart sounds saved the person's life and therefore they partly explain why the person continues to exist, and hence why the heart continues to be

(ii) according to the consequentialist view, the function of the heart is to pump blood, rather than to make noise, because the heart's pumping blood contributes to some important activity of the system within which it is contained, and heart sounds do not. In this case, pumping blood contributes to circulation and this, in turn, to the survival of the organism. This solution corresponds to the view that the *function* of an entity consists in a (special sort of) consequence that it produces, and has nothing to do with the cause or origin of the item itself.

3.1.2 Etiological Theories of Function

There are two main versions of the etiological approach: one which refers to the *reasons* that motivate a purposeful being to create a functional object ("representationalism"),[64] and one that refers to the natural history of the functional entity, independently of the notion of representation. (The latter is typically referred to as "etiological", although "etiological", properly speaking, could refer to either view; in the following the latter will simply be referred to as "non-representational theories of function".) These views will be elaborated in turn.

Representationalist Theories of Function

If one takes an *artifact* as the paradigmatic case for analyzing "function", then the most obvious way for something to come to possess a function is for somebody – an intelligent being – to create it for a purpose. Hence the function of the hammer is to strike nails because striking nails is the purpose people have in mind when they manufacture

there. Does that mean that that person's heart comes to have the function of making throbbing sounds? These sorts of cases will be described in greater detail below.

[64] Although "representationalism" may not seem to be the most accurate title for this view, what is central to the view, as will be suggested below, is that it makes some reference to a prior *representation* (presumably on the part of an intelligent agent) of the functional activity.

them, and that purpose explains "why they are there": that is, why hammers exist, or why they have the form they do (large metal head with firm non-slip handle), or why they are located where they are (in the toolbox next to the nails).[65] This basic observation underlies what one can call "representationalist" theories of function, according to which, if the function of *X* is *Y*, then a representation of *X*'s doing *Y* is part of the cause of *X*'s existence, form or location. This solves the problem of "backwards causation", insofar as the effect does not produce the cause, but a prior *representation* of the effect produces the cause.

To the extent that, in order for a "representation" to exist, it must exist within, or have been created by, a mind, representationalist theories are also *mentalistic* (Bedau [1990]). However, this mentalistic view of function does not appear to be compatible with a modern scientific worldview. This leads to the following question: can there also exist *non-mentalistic* representational theories of function, where representation is analyzed without appeal to minds? Although it may be possible to define a non-mentalistic concept of "representation", this possibility will not be broached any further in this dissertation, mainly because it would involve very similar conceptual issues: for example, should the definition of "representation" be an etiological one, or a consequentialist one?

Non-Representationalist Theories of Function

Whereas representationalist views resolve the problem of backwards causation by seeking the origin of the functional entity in a prior mental representation, non-representationalist views seek to explain why such entities *currently* exist on the basis of entities of the same type that existed in the *past* and that, by virtue of producing the effect

[65] Wright (1973, 158) remarks that the informal locution, "why it is there", captures all of these senses. In Chapter 4, the precise *explanandum* of functional statements will be specified: a function ascription explains the non-zero frequency of an entity within a population (Griffiths [1993, 415]).

in question, were able to persist over time or to reproduce their kind. Hence, the function of an entity is that effect that entities of its kind produced in the past that contributed to the persistence and reproduction of that entity or type of entity. Thus, non-representationalist theories solve the problem of backwards causation by invoking a "cyclical" dimension:[66] *X* did *Y* at time t_0, and as a consequence, *X* persisted until t_1, or *X*, by virtue of doing *Y*, was able to produce entities of the same type as *X* that exist at t_1. Such cyclical modes of reproduction are sometimes referred to as "consequence-etiologies" (Wright [1976, 116]), in that one of the consequences that the functional item produces figures into an etiological account of why it continues to exist at a later time.

The most obvious example of a natural process that generates consequence-etiologies is natural selection, since the differential reproduction of traits that have higher relative fitness than alternative traits explains the maintenance of the former within a population of reproducing entities. To the extent that a trait is selected for, then one can say that the reason it exists at present is that ancestral tokens[67] of that trait produced a consequence in the ancestral environment that bestowed a fitness advantage upon the organisms that possessed it, and (assuming the trait to be heritable) led to its maintenance in the population, thus explaining "why it is there".

Several biologists throughout the twentieth century have drawn attention to the connection between teleological statements and natural selection, and stated explicitly that the existence of natural selection can justify the use of teleology in science.[68] Perhaps the earliest reference comes from the neuroscientist Charles Sherrington, in his *The*

---

[66] It is "cyclical" in the sense that the activity of an entity contributes to its own maintenance and hence allows that entity to continue performing that activity, which continues to contribute to its maintenance, and so on.

[67] A "token" is philosophical jargon for an *instance* of a kind; a "type", for the kind of thing of which it is an instance.

[68] Lennox (1993) argues that Darwin implicitly uses teleological terms such as "end" and "purpose" to refer to the outcome of selection processes (Ibid., 415), though he never explicitly states this fact of his usage.

Integrative Action of the Nervous System (1906). In that work, Sherrington pauses to reflect on his oft-repeated use of teleological terms such as "purpose", and his considerations suggest strongly that he identifies the purpose of a reflex with what it was selected for:

> That a reflex action should exhibit purpose is no longer considered evidence that a psychical process attaches to it; let alone that it represents any dictate of "choice" or "will". In light of the Darwinian theory every reflex *must* be purposive. We here trench upon a kind of teleology…The purpose of a reflex seems as legitimate and urgent an object for natural inquiry as the purpose of the colouring of an insect or a blossom. (Ibid., 235-6)

The ethologist Konrad Lorenz makes a similar remark in his 1963 book, *On Aggression*:

> If we ask "What does a cat have sharp, curved claws for?" and answer simply "To catch mice with," this does not imply a profession of any mythical teleology, but the plain statement that catching mice is the function whose survival value, by the process of natural selection, has bred cats with this particular form of claw. Unless selection is at work, the question "What for?" cannot receive an answer with any real meaning. (Lorenz [1966 (1963), 13-4]; cited in Griffiths [1993, 412])

The evolutionary biologist George Williams also emphasizes this point: "The designation of something as the *means* or *mechanism* for a certain *goal* or *function* or *purpose* will imply that the machinery involved was fashioned by selection for the goal attributed to it" (Williams [1966, 9]).[69]

None of these figures, however, state *why* they believe that explanations based on natural selection fit the pattern of teleological explanations – they simply express, as it

[69] It is ironic that the etiological theory was primarily developed by biologists, since one of the main arguments *against* the etiological analysis is that it does not correspond to actual biological usage!

were, the basic intuition that they *do* without articulating its rationale. Perhaps the first attempt to justify explicitly this view is in the work of the evolutionary biologist Francisco Ayala (1968; 1970) who points out that, in a selectionist explanation, an effect that an entity produces figures into an explanation of why that type of entity currently exists, and this, by definition, constitutes a teleological explanation. As he points out, a teleological explanation is one in which "the presence of an object or a process in a system is explained by exhibiting its connection with a specific state or property of the system to whose existence or maintenance the object or process contributes" (Ayala [1970, 8]). Thus, he draws the conclusion that "the adaptations of organisms…are explained teleologically in that their existence is accounted for in terms of their contribution to the reproductive fitness of the organism" (Ibid., 9). Wimsatt (1972) provides a comprehensive philosophical analysis of the logical structure of function statements and argues that insofar as function statements are construed as teleological explanations, selection processes are the only known and plausible way in which such statements can be justified: "[T]he operation of selection processes is not only *not* special to biology, but appears to be at the core of teleology and purposeful activity wherever they occur" (Ibid.,13).[70] More famously, Wright (1973, 161; also *cf.* Wright [1972]) defines "function" in terms of these consequence-etiologies and argues that natural selection can justify function statements (Ibid., 159).[71]

[70] However, he hesitates to build this insight into a *conceptual analysis* of "function", since he comes up with counter-examples that purport to show that being selected by a selection process is, strictly speaking, neither necessary nor sufficient for having a teleological function (Wimsatt [1972, 15-16]). Moreover, as will be discussed in detail in Section 4.1.5 (under "Wimsatt's (1972) Formulation of SPE"), the concept of selection is comprehensive enough to allow forms of selection other than natural selection operating over an evolutionary time scale, such as immunological selection, synaptic selection, and some forms of learning by positive and negative reinforcement.

[71] Wright (1973), like Wimsatt (1972), does not define "function" explicitly in terms of selection, but claims that having been selected for, in fact, suffices for having a function.

Several different theories of function stem from this basic insight, and much of the philosophical literature on functions consists in the attempt to ramify, extend, and qualify this basic insight. Although it is impossible to exhaustively summarize this literature in a brief space, there are two salient distinctions that can be helpful in mapping the space of non-representational etiological theories (see Table 3.1), thus creating four different etiological explications of, "the function of *X* is *Y*". Furthermore, there are at least two additional variables that one may choose to introduce (that is, in addition to *X* and *Y*), thus generating a potentially limitless number of different etiological theories (see Table 3.2). Each of the two major distinctions will be elaborated in turn; then the two additional variables will be introduced. The breadth of possible theories that can be generated should be apparent.

	Weak Etiological	Strong Etiological
Reproduction-based	Buller (1998; 2002)	Neander (1983; 1991a; 1991b) Millikan (1984; 1989a; 1989b) Brandon (1990) Griffiths (1992; 1993) Godfrey-Smith (1994) Mitchell (1993; 1995) Allen and Bekoff (1995a; 1995b) Schwartz (1999)
Persistence-based	Wright (1973; 1976)	Wimsatt (1972)

Table 3.1: Four types of etiological theory. (See accompanying text for details.)

First Distinction: Weak vs. Strong Etiological Theories

A distinction that will be useful in the dissertation is that between "strong" etiological theories (SE) and "weak" etiological theories (WE), which has been implicit in much of the literature but only clearly articulated by Buller (1998; 2002). According to SE, a function of a trait is an effect that, in the past, the trait was selected for by natural selection. This is the theory of function that was presented earlier, and is sometimes called the "selected effects" theory (e.g., Neander [1991a]). Some version of it is probably the most widely held theory of "function" amongst philosophers (Neander [1983; 1991a; 1991b]; Millikan [1984; 1989a; 1989b; 1993]; Brandon [1990]; Griffiths [1992; 1993]; Godfrey-Smith [1994]; Mitchell [1993; 1995]; Allen and Bekoff [1995a; 1995b]; Schwartz [1999]).

According to WE, the function of a trait is an effect that, in the past, contributed to the reproduction of its bearer and thereby contributed to its own reproduction, *regardless* of whether it was selected for – that is, regardless of whether the requisite variation existed upon which selection could act, or whether existing variation was correlated with differential reproduction. Another way of formulating the distinction is that SE emphasizes the contribution of a trait to *differential* reproduction; WE emphasizes reproduction as such. Both theories, clearly, only ascribe functions to heritable traits.[72]

	Temporal Restriction (*T*): Recent Selection Necessary	No Temporal Restriction
System Restriction (*S*): Selection over Organisms		Neander (1983; 1991a; 1991b)
No System Restriction	Griffiths (1992; 1993) Godfrey-Smith (1994)	Millikan (1984; 1989a; 1989b) Brandon (1990) Mitchell (1993, 252; 1995) Schwartz (1999)

Table 3.2: Addition of system and temporal variables to strong reproduction-based etiological theories. (See accompanying text for details.)

A simple example drawn from Dover (2000, 41) can help to clarify the distinction. Suppose that, in a small population, genetic drift carries an allele to fixation at t_0. Although that allele has a phenotypic effect, it did not confer any fitness advantage on its possessors. Now suppose that, at t_1, the environment changes in such a way that

[72] Buller (2002, 230-33) points out that it is not uncommon for philosophers to vacillate between SE and WE.

possession of the allele becomes necessary for survival. Even though all of the individuals within the population have the allele – so there is no selection for it – they all would have perished at t_1 had any of the alternate alleles available at t_0 gone to fixation. Thus, at t_2, it can be said that one of the reasons that the allele currently exists is because it produces the effect in question. In this sense the scenario satisfies the pattern of teleological explanation. But since selection did not enter the scenario, SE does not bestow a function upon the trait, since at t_1, the requisite variation did not exist upon which selection could act, and at t_0, the differential reproduction of alleles was not correlated with differential fitness. Hence, the weak etiological theory is clearly more liberal with respect to the range of evolutionary mechanisms that it considers relevant to function ascriptions, yet it still permits teleological explanation. Schlosser (1998, 323) and Sarkar (2005, 18), for example, argue that selectively-neutral traits should be able to have functions, especially since they are capable of playing important roles in survival and reproduction.

Second Distinction: Reproduction-based vs. Persistence-based Theories

This second distinction, like the first, is implicit in much of the literature although attention is rarely brought to it. It is typically assumed that in order for a *part* of a system to have an etiological function, that type of part must have contributed to the persistence or reproduction of ancestral systems, and thereby contributed to its own intergenerational reproduction via the mechanisms of heredity. This assumption is virtually taken for granted in much of the literature when SE is adopted, since it is often accepted that by definition, natural selection involves the differential *reproduction* of traits (e.g., Lewontin [1970, 1]). Hence, according to *reproduction-based* accounts, the function of an entity consists in doing whatever it was that prior tokens of that same type of entity did that led

to their reproduction (typically by contributing the persistence or reproduction of the containing system).

According to *persistence-based accounts*, the function of a given token of an entity consists in doing whatever it was that that token did that contributed to its persistence over time, whether or not it contributed to its intergenerational reproduction. Many biological traits contribute to their own persistence; in fact, any trait that contributes to the survival of the trait-bearer, where the survival of the trait-bearer is necessary for its own continued existence, fits this pattern. Thus, to take an example from Wimsatt (1972, 43; see *fn*. 64 of that paper), the heart beats; in beating it circulates the blood, which strengthens, among other things, the rib cage, which protects the heart. Thus the heart, by beating, actively contributes to its own persistence over the lifespan of the individual.[73] The notion that in order to have a function, a part need only contribute to its own (intragenerational) persistence, and not necessarily to its (intergenerational) reproduction, is central to McLaughlin's (2001) theory of function. Dretske (1988, 98-101), as well, suggests that an entity can come to possess a function by doing something that leads to its "recruitment" within a system, even if it is the only instance of its kind and is not hereditary. [74]

---

[73] Certainly, one could say that a non-beating heart also persists (Sarkar, pers. comm.). Yet one may reasonably withhold the function of persisting to a non-beating heart because, in a sense, it is not "doing" anything to contribute to this persistence. Contrast this with the cells that make up the non-beating heart, which "do" something to contribute to their own persistence, insofar as the metabolic processes they carry out ensure their continuation, at least for a short time, after the heart stops beating.

[74] In the following, the expression "persistence-based accounts" will be used broadly to include *reproduction-based accounts* as well. Consequently, according to a persistence-based view, in order for an entity to have a function it must have contributed to its own *persistence or reproduction*. This is useful because it permits the persistence-based theory to assign functions to entities or processes (such as childbirth) that decrease the chances that an individual organism will persist while increasing the chances that it will reproduce. Sarkar (2005, 18) introduces a similar distinction with the concepts of "broad-sense" and "narrow-sense" function (see Section 3.1.2, under "Fitness-Contribution Theories"). Something has a broad-sense function if it contributes to the persistence of the system in which it is contained, and a narrow-sense function if it contributes to the fitness of that system.

The two distinctions that have been introduced – that between strong and weak etiological theories, and that between the reproduction-based and persistence-based theories, crosscut one another, creating four different theories. (See Table 3.1.) A *strong reproduction-based etiological theory* (SRE) is one according to which, in order for an entity to have a function, it must have been selected for by natural selection operating over a population of reproducing entities (it must have undergone differential reproduction). A *strong persistence-based etiological theory* (SPE) is one according to which, in order for an entity to have a function, it must undergo differential persistence, rather than differential reproduction. Dawkins (1989 [1976], 12) discusses such cases under the rubric of "survival of the stable" rather than "survival of the fittest": for example, atoms that tend to fall into stable patterns tend to outlast those that do not (Ibid., 13). Hence, this constitutes a form of selection that operates over molecular structures. In other words, selection does not only operate over populations of reproducing entities, but over entities that, by virtue of differences in their ability to utilize environmental resources, persist for different periods of time.

It might be thought that differentially persisting entities that are not capable of some form of reproduction would be fairly *biologically* uninteresting because (so the argument might go) they are incapable of giving rise to complex adaptive structures (as natural selection operating over reproducing entities does), and hence biological "functions" should not be assigned to such entities. Yet one paradigmatic example of such a complex adaptive system – the mature synaptic structure of the brain – is partly due to selection processes that operate over neural projections that are themselves incapable of reproduction. Synaptic structure is often the result of *differential retention or amplification* of different synapses on the same target neuron, instead of their differential

reproduction (Changeux and Danchin [1976]; Edelman [1987]).[75] Hence, SPE assigns functions to unique, non-hereditary structures, and SRE does not, or at least not without some refinement. *Neural selection* processes will be described in more detail in Chapter 4.[76]

A *weak reproduction-based etiological theory* (WRE) is one according to which, in order for an entity to have a function, ancestral tokens of that entity must have done something that contributed to their own reproduction, even if they were not selected for. Buller (1998; 2002), as described above, presents such a theory. A *weak persistence-based etiological theory* (WPE) is simply one according to which something comes to have a function because, in the past, it did something that contributed to its own persistence over time: each individual's heart, as pointed out above, satisfies this ascription because it contributes to the survival of the system which it in turn depends upon for its continued persistence. Wright's (1973) explicit definition of function qualifies as a WPE account (see Chapter 4.1.2 for the shortcomings of this account). These four theories will be evaluated in Chapter 4, and the choice of a *strong persistence-based* theory of function will be defended (Section 4.1.5).

Two Additional Variables: System and Temporal Variables

Note that, although all four categories afforded by the rudimentary taxonomy of Table 3.1 are filled, there is a significant imbalance in the existing literature, in that the majority of attempts fall under SRE. Nonetheless, there are significant discrepancies

[75] Also see Darden and Cain (1989), who formalize a concept of "selection" which is more general than that presented in Lewontin (1970) in order to permit such phenomena to qualify as undergoing "natural selection". However, Darden and Cain's analysis purchases this generality at the price of invoking the problematic concepts of "benefit" and "suffering" (Ibid., 116). See pp. 122-23 for their application to Edelman's theory of neuronal group selection.

[76] Wimsatt (1972, 14-15) clearly intends this general notion of selection in his theory of function, since he argues that any explanation that appeals to "blind variation and selective retention" is teleological, and he regards trial-and-error learning procedures as a type of selection (see Section 4.1.5, under "Wimsatt's (1972) Formulation of SPE").

between these attempts. This suggests that the taxonomy presented is too coarse-grained, and a more fine-grained classification would be useful in articulating the space of etiological theories further. This space can be sufficiently articulated for the present purpose by relativizing function ascriptions to two additional variables. Until this point, it has been assumed that there were only two relevant variables in the *definiendum*,[77] "The (or *a*) function of *X* is *Y*" – namely, the entity that possesses the function, *X*, and the functional effect that it produces, *Y*. It was assumed that all other variables could be left implicit. Two additional variables will be introduced, a *system* variable, *S*, and a *temporal* variable, *T* (see Table 3.2).[78]

The first variable that will be introduced, *S*, describes the *system* of which the functional entity is a part, thus yielding the "elementary sentence" for function that will be analyzed here: "The function of *X* in *S* is *Y*". There are three reasons for relativizing the function ascription to the system of which the entity is a part. The first is that one fairly common intuition about functions – at least in the case of natural functions – is that only *parts* of systems can have functions, and not systems taken as *wholes*.[79] The human

---

[77] "Definiendum" refers to the term to be defined; "definiens" the expression used to define it. Compare "explanandum", the phenomenon to be explained, and "explanans", the theory doing the explaining.

[78] Wimsatt (1972, 32) presents a much more structured definition, where the function of an entity is explicitly relativized to a *behavior* of that entity, a *system*, an *environment*, a *purpose*, and a *theory*. The reason he incorporates all of these additional variables is that he wants the function of an entity to be *uniquely* determined once those variables are specified (Ibid., 33). However, in this dissertation there is no insistence that the function of an entity must be uniquely determined (see Section 5.1, which accepts a fundamental indeterminacy in function ascriptions). Hence many of these additional variables will not be explicitly incorporated into the definition of "function" presented here.

[79] Wright (1973, 145) objects to this relativization of function to *systems*, primarily on the basis of his consideration of artifact functions. A watch, for example, has a function, but it is not clear that this function is relative to any system of which it is a *part*. One should say, rather, that the watch has a function *for* a system – namely, the person who uses it or benefits from it. But this introduces new and potentially troublesome concepts, namely, the idea of *functioning for* and the idea of *benefit*. Perhaps one could extend the concept of "system" to allow the *person wearing a watch* to constitute the relevant system of which the watch is a *part*. Nissen (1997, 37), however, argues that such a move generates bizarre counterexamples, since it imposes no principled restrictions on what constitutes a "system": for example, Mr. Smith can use a rock as a paperweight, but it seems bizarre to relativize the function of the rock to the "Smith-rock" system of which it is a part. Perhaps Godfrey-Smith's (1994, 349) insistence that (biological) functions can only be ascribed to parts of "biologically real systems" (see below) can help to impose such a restriction. On the other hand, it is arguable that functions are not always ascribed to "parts" of systems that they subserve

digestive system has a *function* – to absorb nutrients and eliminate wastes – but humans as such, considered as autonomous units, do not have functions. However, if a human being is conceived of as occupying a place within a larger system, such as an ecosystem, then he or she may be said to have a function – e.g., the production of soil nutrients and carbon dioxide. The function of an entity, then, and even whether or not it can be said to *have* a function, seems to depend crucially on the sort of system of which the entity is considered a part. Thus, this dependence should be reflected by the introduction of an additional variable, *S*, to represent the system of which the item is a part – or at least, of the system which the item was once, in the past, a part.[80]

A second reason that functions should be relativized to systems becomes evident if one adopts some version of the strong etiological (SE) theory, since selection involves the propensity of a trait to contribute to the differential survival or reproduction of an inclusive system, as a consequence of which it, being heritable, ensures its representation in future generations. Moreover, even though WRE does not appeal to *selection*, nonetheless, it appeals to a trait's contribution to the reproduction of some inclusive system, and hence it retains the reference to the system of which the entity is a part.[81]

The third, most important reason that the etiological theory should relativize function ascriptions to systems is that there is no *unique* type of system that a trait must contribute to in order to be maintained within a population.[82] A nucleotide segment, for example, can be maintained by natural selection because it contributes to the fitness of the chromosome in which it resides, the individual organism in which the chromosome

---

even in the biological realm. For example, the galls that grow on oak trees have a function for the gall wasps that produce them, and not for the oak tree (Griffiths [1993, 416]).

[80] This qualification would allow, e.g., transplanted organs to retain their functions after being removed from the donor.

[81] Recall that "part" is here being used in a sufficiently broad sense to include not only spatio-temporally discrete objects but also activities and qualities.

[82] Obviously if there were such a unique system one would not need to relativize the function ascription to a variable, but simply incorporate a reference to that system within the definition.

resides, or that organism's group. Thus selection operates at several levels (Lewontin [1970]). More importantly, the same trait may make conflicting contributions at different levels. A central example of such conflict is meiotic drive (or segregation distortion), which refers to any process that causes certain alleles to be overrepresented in the sex cells. Chromosomes that carry segregation distorter (SD) genes appear to operate by doing something to sabotage the homologous chromosome, disrupting normal sperm development for the homologue (Crow [1979, 138]). Consequently, segregation distortion increases the fitness of the SD chromosome, although it can decrease the fitness of the organism. In *Drosophilia melanogaster*, for example, males that are heterozygous for SD have fewer functional sperm than those that do not carry the SD chromosome, and under certain experimental conditions this can lower the fly's fitness (Ibid., 137); those which are homozygous for the SD genes are sterile.

Similarly, Darwin speculated that sterility in worker ants, while decreasing the fitness of sterile individuals, could increase the fitness of the group to which their labor contributes:

> [S]ome insects and other articulate animals in a state of nature occasionally become sterile; and if such insects had been social, and it had been profitable to the community that a number should have been annually born capable of work, but incapable of procreation, I can see no especial difficulty in this having been effected by through natural selection. (Darwin [1998 (1859), 352])[83]

When selection processes are described in the context of evolutionary biology, it is often assumed that the relevant level of selection is the *organism*. This familiarity has led some philosophers to explicitly restrict the *type* of system over which the function-

[83] The explicit specification of the system-level over which selection processes are operating will be crucial in the following chapter, which describes neural selection theories, in which selection operates at the level of *synapses* (or more accurately, neural projections), *neurons*, and *groups of neurons*, and has important implications for the formation of mature synaptic networks.

bestowing selection process can act to the *organism*. For example, Neander (1991a) defines "function" in the following terms:

> It is the/a proper function of an item (*X*) of an organism (*O*) to do that which items of *X*'s type did to contribute to the inclusive fitness of *O*'s ancestors, and which caused the genotype, of which *X* is the phenotypic expression, to be selected by natural selection. (Ibid., 174)

Buller (1998) incorporates a similar restriction into his initial formulation of WRE:

> A current token of a trait *T* in an organism *O* has the function of producing an effect of type *E* just in case past tokens of *T* contributed to the fitness of *O*'s ancestors by producing *E*, and thereby causally contributed to the reproduction of *T*s in *O*'s lineage. (Ibid., 507)[84]

The analyses offered by Brandon (1990, 192-3), Griffiths (1993), and Godfrey-Smith (1994) explicitly eschew such a restriction on the type of system over which the function-bestowing selection process can act. Recognizing the plurality of such levels, Godfrey-Smith incorporates into his analysis the qualification that the functional entity must reside within a larger "biologically real system" (Ibid., 349) and that the entity must have been selected for due to a positive contribution to the fitness of this system (also *cf*. Griffiths [1993, 416]).

The second variable that function statements will be relativized to is the *temporal* variable. If a trait has a function because of the fitness contribution it made in the past, then how recently in the past must it have so contributed in order to retain its function? Should the function ascription be relativized to some specific time span, *T*?

[84] Later in the same article, however, he points out that this restriction to organisms just represents the paradigm case, and that "the weak theory [i.e., WRE] can attribute functions within any type of system that biological theory finds it necessary to represent as possessing fitness" (Ibid., 516).

Perhaps the most obvious motivation for imposing such a restriction is to allow for the possibility of vestiges, what are often said to be *functionless*. But if they are adaptations, then their past contribution to the fitness of ancestral systems figures into a complete explanation of why they still have a non-zero frequency in the present population. Therefore, without imposing any temporal restrictions on the explication of "function", it is not clear how that explication can capture the idea that a heritable trait, though it once possessed a function, no longer does, but has been retained because the relevant mutations that would have allowed it to atrophy or be replaced never arose. The rudimentary ocular cyst of the cave-dwelling fish, *Phreatichthys andruzzii*, is not a dysfunctional eye, but a functionless vestige – even though at some point the organ had been selected for because of sight.[85] This suggests that the elementary statement of the function ascription should read, "The function of *X* in *S* relative to *T* is *Y*".

Another case which supports the need for introducing temporal restrictions on function ascriptions is the case of functional co-option, in which a trait that initially spread within a population by selection for one of its consequences eventually came to be maintained by selection for something else, or in which a trait that was initially not selected for at all came to be selected for in a new environment – such as feathers, which were initially selected for because of insulation and only later because of flight. Gould and Vrba (1982) introduce the well-known distinction between adaptation and exaptation, which partly overlaps the distinction made above. A trait is an *adaptation* if it was "built by selection for its current role" (Ibid., 6), and an *exaptation* if it was later "co-opted" for a useful role that it was not originally selected for. Such cases are ubiquitous in the biological world and render problematic any simplistic attempt to infer the selective

[85] Interestingly, *Phreatichthys andruzzii* is not actually anopthalmic (born without eyes); rather, eye development begins very shortly after egg laying, and eye degeneration occurs after about 36 hours and is complete within one month (Berti *et al*. [2001]).

history of a trait from its current contribution to fitness. Any etiological theory must possess the resources to conceptualize such transitions appropriately.

There are three main approaches to instantiating the temporal variable on function ascriptions. One approach requires that, in addition to its past activity, a functional trait must *presently* be capable of performing its function. A second approach requires that, in order to have a function, the trait must have contributed to its own reproduction in the *recent evolutionary past*, even if it is not currently capable of performing that function. A third approach requires that, in order to have a function, the trait must have contributed to its initial spread within a population, regardless of what it did to contribute to its recent maintenance and regardless of its present activity. This approach would relativize the function ascription to the *early history* of the trait, that is, the time period shortly following its introduction into a population. Each of these three approaches will be described in turn.

One fairly obvious way of resolving the problem of vestiges is to restrict function ascriptions to those traits that are *presently* capable of producing the same effect that, in the past, made a positive contribution to fitness. Wright (1973; 1976) presents such a solution by formulating a "tenseless" explication of function that, in its generality, implies that the part in question is still capable of performing the function for which it was once selected. (Wright's definition of "function" will be evaluated in Chapter 4.1.2. It constitutes a version of WPE.) According to his explication, "the function of *X* is *Y*" means that, "*X* is there because it does *Y*" and that "*Y* is a consequence of *X*" (see Wright [1973, 161; 1976, 81]). Clearly, the purpose of Wright's fairly general analysis is to present the idea of a "cyclical" causal process, in the sense specified above (see *fn*. 66). The statement that "*X* is there because it does *Y*" in the first condition refers, as it were, to a universal generalization about things of type *X*, namely, that they are capable of doing

Y, regardless of time, place, or circumstance. In other words, the implication is that *X* is there (presently) because *things of type X are capable of doing Y* (as a universal generalization). This has the consequence that if something was initially selected for because of some effect but is no longer capable of producing it, then it would not have that function because that statement would be false of it. Thus, among other things, his explication of function ascribes to *X* the current *capacity* of doing *Y*. Hence, according to the use of "etiological" and "consequentialist" defined above, Wright's theory, strictly speaking, is a consequentialist one, because it entails that *X* must currently be capable of contributing to *Y*.

Kitcher (1993) presents a similar restriction. In order for a trait to have a function, not only must selection be partly responsible for maintaining the trait in the recent past, but it must continue to do so: "The function of *X* is *Y* only if selection of *Y* is responsible for maintaining *X* both in the recent past and in the present" (Ibid., 387). He argues that this restriction allows a theory of function to enjoy the benefits associated with *both* the etiological approach and the consequentialist approach, since it allows function statements to perform a dual role: they can explain why an entity has been maintained in a population, and they sketch a prediction about how that entity will continue to be maintained in that population in the future (Ibid., 386).[86]

[86] Unfortunately, by combining both views his theory enjoys the benefits of neither – at least relative to the narrowly-defined purpose of this dissertation. The reason is the following. One major argument for *consequentialist* theories of function is that they are held to be in better accord with actual biological usage (see Section 3.1.3). When evaluating the function of an entity, practicing biologists primarily look towards the current capacities of that entity, and, more specifically, its current role in contributing to fitness, rather than its historical contribution. But this admitted benefit of consequentialist theories is lost if one insists that function ascriptions incorporate some etiological component, since it entails that standard biological practice does not suffice to warrant function ascriptions. Similarly, one major argument for *etiological* theories of function is that they make sense of the *normative* component that function ascriptions often carry – that is, if something has a function it can fail to perform that function. But by insisting that if an entity has a function then it must *currently* contribute to fitness disallows such a normative component, or at least, it is not normative in the sense that it satisfies both adequacy conditions CA_1 and CA_2* outlined in Chapter 2.2. (The claim that consequentialist theories cannot be normative in this sense will be defended in Section 3.2.1.) Walsh (1996, 564; see *fn*. 15 of that paper) makes a similar point about how Kitcher's (1993) "mixed" approach loses benefits associated with either.

A widely accepted approach amongst philosophers of biology is that which identifies the function of a trait with what it was selected for in the *recent evolutionary past* (Griffiths [1992, 1993]; Godfrey-Smith [1994]), even if that trait is currently unable to perform the function. But how should such a temporal unit be estimated? For it must specify a unit of time that allows one to decide how long a trait can be maintained within a population without contributing to its own maintenance before becoming a vestige. Griffiths (1992) defends a version of SRE according to which the trait in question must have contributed to its maintenance in a population during the last "evolutionary significant time period" for that trait, and he defines an evolutionarily significant time period for a trait, *T*, as a time period during which, given the mutation rate at the loci controlling *T* and the population size, one would have expected some regression (atrophy) of *T* were it not making some contribution to fitness (Ibid., 128). Godfrey-Smith (1994), while introducing the expression "modern history theory of functions", leaves the determination of such a unit implicit.

Godfrey-Smith (1994) argues that the modern history approach, in addition to resolving the problem of vestiges, can help to bridge the apparent gap between etiological theories of function and actual biological usage (see Section 3.1.3). As he points out (Ibid., 351), according to an influential set of distinctions introduced by Tinbergen (1963), the field of behavioral ethology is largely concerned with four questions concerning behavior: its (proximate) causation, its survival value, its evolution, and its ontogeny (Ibid., 411). In Tinbergen's usage, "survival value" is synonymous with "function", and explicitly separated from the question of evolution, and in particular, from the selective history of a behavior (Ibid., 423). This suggests that in behavioral ethology, "function" is often taken to be a purely consequentialist notion (but see the quote from Lorenz [1966 (1963)] above, which suggests a pluralism of function concepts

within ethology). Although the modern history approach does not, of course, allow "function" to be severed from "history", it does allow the preservation of a four-part distinction which is created by dividing the category of "evolution" into two parts: that concerning the *initial spread* of the trait within the population, and that concerning its *recent maintenance* in the population (Godfrey-Smith [1994, 356]). If one assumes a correlation between current survival value and recent selection, then a rough extensional correspondence between the two classifications could be maintained.

However, Schwartz (1999) questions whether such a correlation can be assumed. Certainly, if something currently possesses survival value, then unless it is a novel trait, one can safely assume that it did something in the recent past that contributed to the reproduction of ancestral organisms and hence helps explain its maintenance in the population. But that does not provide reason to believe that it underwent selection in the recent past. Consequently, Schwartz relaxes this restriction on the modern history view of functions, arguing that if a trait was selected for at *any time* in the past because it did *F*, and if it recently *contributed to survival or reproduction* by doing *F*, then it should be said to have the function *F*, even if it did not *undergo selection* for *F* in the recent past (Ibid., S219).

An alternative to restricting the function of an entity to what it was selected for in *recent history* would be to restrict the function of an entity to what it was *originally* selected for, that is, to whatever explains its initial spread within a population (assuming that natural selection accounts for its initial spread). In this case, the temporal variable would be indexed to the early history of the trait. This is the option explored by Gould and Vrba (1982). Their argument begins by defining "function" in terms of "adaptation" (Ibid., 5), and then defining "adaptation" in a very restricted sense to describe what a trait was selected for upon its initial appearance (as opposed to "exaptation", which describes

how traits are later co-opted for different uses) (Ibid., 6).[87] However, although this distinction between "adaptation" and "exaptation" may be useful,[88] it is not clear what motivates their narrow restriction of functions to adaptations, especially because it leads, by their admission, to claims that seem so contrary to typical biological usage – for example, that the function of feathers is not flight, since they appear to have originally been selected for because of insulation (Ibid., 7).[89]

With the introduction of the system variable, *S*, and the temporal variable, *T*, a more refined classification of SRE theories can be provided (Table 3.2). In particular, the introduction of those variables helps to articulate the space of those theories that appeal to natural selection operating over reproducing populations – although, in principle, analogous distinctions could be raised for any version of the etiological theory of functions.

3.1.3 Consequentialist Theories of Function

Despite the plurality of etiological theories, and despite the attempts to render etiological theories more consistent with modern biological usage, it is often pointed out that typically, when biologists seek to determine the function of an entity, they look to some subset of current dispositions or capacities of the entity rather than to the fossil record (Amundson and Lauder [1994]; Godfrey-Smith [1994, 351]; Walsh [1996, 558];

---

[87] Gould and Vrba use the term "aptation" to be neutral between an "adaptation" and "exaptation" (Ibid.).
[88] Though see Reeve and Sherman (1993), who question its utility in the practice of evolutionary biology.
[89] Allen and Bekoff (1995a), however, argue that the distinction between "adaptation" and "exaptation" conceals a very important teleological distinction, namely that between "function" and "design". Unlike the concept of function, which can be used broadly to encompass whatever a trait was selected for, the concept of design should only be applied to that subset of functions that partly explain the structural modification of a trait over time (1995a, 615). They point out that, often, what something is an "adaptation" for (in Gould and Vrba's sense) is often what it is "designed" for, and that "exaptations" will often correspond to traits which merely have "functions" but were not designed, since they did not undergo any structural modification. Buller (2002), however, argues that the distinction between "function" and "design" is unprincipled, because whether something is *designed for X*, or merely has the *function of performing X*, often depends upon purely conventional decisions about how selection pressures should be individuated.

Schlosser [1998, 304]; Wouters [2003, 658]; Sarkar [2005, 18]; Griffiths [2005]). Although, as noted in Section 3.1.2, biologists sometimes *do* use "function" more or less synonymously with "adaptation", in many contexts "function" is tied more closely to the current survival value of a trait, regardless of its origin. For example, as noted above, Tinbergen (1963) equates "function" with "survival value", and explicitly separates questions of function from those of evolution (Ibid., 423). Mayr (1961), similarly, distinguishes "functional biology" and "evolutionary biology", arguing that the former is concerned with the realm of "proximate causes" and the latter, "ultimate causes" of an entity or process, whereas, according to etiological theories, "function" typically describes the realm of ultimate causes. For Mayr, "functional biology" primarily describes "how" something works and not "why" it is there, although the scientific utility of any absolute distinction between "how" and "why" questions has been criticized (Sarkar [2005, 19-20]).

Even more broadly, "function" is often used to characterize the entire range of activities that a part of a system is capable of performing (e.g., the sense in which "function" is opposed to *structure*). For example, the evolutionary morphologists Bock and Von Walhert (1965) define the function of an entity simply as "all physical and chemical properties arising from its form" (Ibid., 274), provided that these properties are not relative to the environment. Amundson and Lauder (1994) argue that this more liberal usage is standard in anatomy, comparative morphology, and physiology. This makes the use of function statements in those disciplines heavily dependent upon the interests of the investigator, since without at least imposing a pragmatic restriction on the appropriate use of function statements, virtually every structure in the natural world can be said to possess a "function".

These examples suggest that despite the modifications that can be imposed on the etiological theory, it does not adequately capture the majority of biological usage.[90] Thus, some argue, functions, whatever else they may be, must be thought of as current dispositions of traits, and if this is incompatible with the view that present-tense function ascriptions constitute explanations for the current presence of a trait in a population, then one must reject the explanatory status of function ascriptions.

As noted above, consequentialist theories of function almost invariably conceive of the function of an entity as consisting in its contribution to something else. Insofar as functions, in the biological context, are typically ascribed only to *parts* of systems (rather than the system as a whole), then, according to consequentialist theories, the function of an entity consists in its contribution to some property or capacity of a more inclusive system – e.g., the contribution of a trait to the fitness of the organism. Hence, in the following, consequentialist theories will be classified in terms of the *sort* of system property or capacity which performance of the function contributes to.[91] Four such

[90] It is peculiar that Neander (1991a) defends SRE as an accurate conceptual analysis of modern biological usage, given that it is often argued that it fails precisely as such an explication. Her article does not respond to that challenge. However, whether that challenge is ultimately successful depends on what, precisely, an explication of "function" is intended to accomplish, and on whether a discernable plurality of function concepts exists even within biological usage. Millikan (1989b), for example, argues that her explication is not intended as a *conceptual analysis* of modern biological usage, but as a *theoretical definition*, in the same sense in which "being H_2O" constitutes a theoretical definition of "water". Schwartz (2004) goes further by emphasizing the stipulative and constructive roles of philosophical definitions of "function", arguing that such definitions constitute *explications* of biological usage, rather than conceptual analyses or theoretical definitions. According to Carnap (1950, see Chapters 1 and 2), philosophical explication involves the replacement of a vague concept by a precise one, and hence it often entails making distinctions that did not previously exist in the scientific context in question. It has the character of a proposal, to be accepted or rejected on pragmatic grounds. As stated in the introduction to this chapter, this dissertation adopts a pluralistic view, in that the concept of function it proposes is only intended to be appropriate for the biological perspective in psychiatry (its explication being guided by the two adequacy conditions outlined in Section 2.2), regardless of whether it is appropriate for other biological usages.

[91] It is not always the case that consequentialist theories define the function of an entity in terms of its contribution to something else. As noted above, according to Bock and von Wahlert's (1965) liberal conception, the function of a structure consists of more or less the totality of effects produced by its structure. In this theory there is no sense in which a function contributes to anything else, much less a containing system. By the same token, it is not always the case that, when a functional entity does contribute to a system, that system is its own inclusive system. This is most obviously true in the case of artifacts (see *fn*. 79).

theories will be presented: *interest-contribution* theories, *goal-contribution* theories, *good-contribution* theories, and *fitness-contribution* theories. These are all separable in principle, although in practice they may identify the same "functions". Moreover, this list is not intended to be exhaustive – certainly, one can imagine other such theories, or variations on those listed. However, this classification fairly adequately spans the range of those that have actually been proposed.

Unlike the case of etiological theories, there are very few salient distinctions with which to articulate the space of consequentialist theories in any principled way. One salient distinction would classify consequentialist theories into *substantive* and *methodological* approaches. A substantive approach to function is one that is essentially concerned with a problem of *demarcation*. There are two levels to this demarcation problem. The *first-level* demarcation problem is that of distinguishing between the sorts of *systems* the parts of which can have functions from those that cannot. Why are functions typically assigned to the parts of an organism, and not to a pile of rocks? Once such a system is selected, the *second-level* demarcation problem concerns assigning functions to certain effects of a trait, and not others. Why does the kidney have the function of extracting waste from the blood, rather than supporting hard calcium formations along its inner wall? Clearly, strong etiological theories of function are substantive theories, since they require any system that can have functions to be capable of undergoing selection (first-level demarcation), and they distinguish a function of a trait from a mere effect in terms of its role in that selection process (second-level demarcation).

In contrast, a methodological approach to function is one that elaborates the distinct manner in which scientists theorize about, analyze, or explain systems that are believed to have functions. They are concerned more with a style of analysis that might

be called "functional analysis" (e.g., Hempel [1965 (1959)]; Cummins [1975; 1983]) (although one that should not be confused with a branch of mathematical physics that goes by the same name!) "Methodological" questions will be characterized in more detail below.

In the following, what will be referred to as *interest-contribution* theories constitute methodological approaches to function, because the activities that are considered to possess functions are determined by the interest of an investigator and hence by their relevance to a *specific explanatory context*. Virtually no substantive constraints are imposed upon what sort of entity can have a "function". The remainder – *goal-contribution*, *good-contribution*, and *fitness-contribution* theories – constitute substantive approaches, because their primary objective is one of demarcation.[92] This list is not intended to be exhaustive, but representative of the range and variety of contribution-based theories that are available.

Methodological Approaches to Function

Methodological features of functional explanation in biology have not received nearly the attention by philosophers of science that substantive features have. For example, how do scientists identify the relevant "parts" of a system when assigning functions to them, and in what way is this identification guided by prior assumptions about what the system (as a whole) is doing (Kauffman [1970])? Functional analysis often depicts the relevant activities of the system hierarchically: various subsystems contribute to the capacities of more inclusive systems, which, in turn, contribute to the capacities of even more inclusive systems. What sorts of assumptions govern the

[92] The claim that *goal-contribution* theories are "substantive" in this sense is controversial; because one of the main problems that such theories have confronted is the problem of vacuousness – namely, that all systems can be seen as goal-directed and hence it does not actually achieve any demarcation (see below, under "Goal-Contribution Theories").

construction of such hierarchies (Craver [2001]; Wimsatt [2002])? Clearly, functional analysis is concerned with systems that can be said to have something like "modular" design. If so, what concept of modularity is operative? (Cummins [2002, 158-9] points out that functional analysis presupposes that the parts of the system in question can, in principle, be replaced by functional equivalents, and this seems tantamount to imputing a type of modularity to that system.)

These sorts of questions illustrate the point that methodological and substantive questions are not entirely separable, since the methodological assumptions inherent in functional analysis place restrictions upon the sorts of systems that can have functions or that are appropriate subjects of functional analysis. For example, an object with a fairly homogenous constitution – that is, with little internal differentiation – could not satisfy any assumptions about modularity or hierarchical organization and hence it would be inappropriate to assign functions to its parts. (However, it could be said to have a function as a part of a larger system.) Similarly, according to Cummins' (1975; 1983) usage of "functional analysis", phenomena that can be explained by straightforward subsumption under a physical law, such as the capacity of a falling body to accelerate, or that of a submerged object to displace water, are not appropriate subjects of functional analysis. Therefore, these methodological assumptions already entail some substantive commitments, at least in terms of the first-level demarcation problem.

Interest-Contribution Theories

The most general contribution-based theory is the interest-contribution view, according to which the function of an entity consists, roughly, in its contribution to maintaining some property of a system that is of interest to an investigator or to a group of investigators. Put simply, to say that, "the function of polar bears' dense, water-

repellent fur is to retain heat" implies that heat retention is a capacity that the speaker or investigator or group is interested in, and the dense, water-repellent fur is capable of contributing to heat retention. In this sense, there is nothing mysterious about function ascriptions, since they do not imply that heat retention explains the presence of fur; rather, they merely imply the commonplace fact that the presence of fur explains heat retention. Another way of stating the idea behind the interest-based approach is that some reference to the interest of the investigator is necessary for establishing the truth-conditions of any particular function ascription.

The most well-known proponent of this theory is Cummins (1975; also see 1983; 2002) – so well-known, in fact, that such functions are often simply referred to as "Cummins functions" (Millikan [1989a]; Godfrey-Smith [1993]), or even "C-functions" (Walsh and Ariew [1996]). Because Cummins' view explicitly relativizes the appropriateness of function ascriptions to their role in a specific explanatory context, it can be called a methodological approach. Similar views that emphasize the explanatory or pragmatic context of function ascriptions are held by Hempel (1965 [1959]), Lehman (1965), Prior (1985), Amundson and Lauder (1994), Hardcastle (1999; 2002), Davies (2001), and Craver (2001).

Because of the fact that, according to these views, functions are only limited by the interests – epistemic or pragmatic – of the investigator, they are often accused of overbreadth. On the one hand, "functions" could be ascribed throughout the non-organic world. For example, a particular arrangement of rocks can have the "function" of contributing the widening of a river delta downstream from it (Kitcher [1993, 390]), and clouds can have the function of promoting vegetation growth (Millikan [1989b, 294]). On the other hand, functions can be applied to entities that are clearly malfunctioning or maladaptive; as Cummins himself points out, the appendix keeps people vulnerable to

appendicitis but it sounds strange to call this one of its functions (Cummins [1975, 752]) – even though medical researchers are clearly interested in providing an analytical account of how this takes place!

Of course, one may wish to accept a fairly liberal construal of "function", according to which the parts of molecules, cloud formations, or volcanoes have the "functions" of contributing to some natural processes (e.g., Amundson and Lauder [1994, 346-7; see Section 4.1). Some scientists, in fact, may find nothing particularly counterintuitive about this usage (Sarkar [pers. comm.]). However, the relevant question in this dissertation is not whether there exists some concept of function according to which clouds can legitimately have the "function" of promoting vegetation growth – as noted in the introduction, this dissertation adopts a pluralistic standpoint on concepts of function. The real question is whether such a liberal construal of "function" is appropriate to the context of psychiatry, where the concept of function must support a corresponding concept of "dysfunction". From this perspective, it seems counterintuitive, or at least highly anthropocentric, to say that clouds are "dysfunctional" when they do not contribute to vegetation growth. This point will be raised again in the next chapter (see Section 4.1.2., under "Intuitive Implausibility of WPE").

Substantive Approaches to Function

As noted above, goal-contribution, good-contribution, and fitness-contribution theories are "substantive" in that they seek to delineate a certain type of system that is an appropriate subject of functional analysis (first-level demarcation) and they seek to justify why certain ascriptions of functions to its parts are correct, and others incorrect (second-level demarcation). All three substantive contribution theories entail, for example, that a stick pinned to a rock and stuck in place by its own backwash does not have a "function", but they disagree as to why this is so: a stick being pinned to a rock

does not contribute to any system goals; it does not contribute to the good of anything; or it does not contribute to the fitness of any system. Similarly, they all entail that the function of the kidney is to extract water from the blood, but disagree as to why this is so: because that is the kidney's specific contribution to the goal of the water-regulatory system; because it contributes to the organism's good; or because it contributes to the organism's fitness. Certainly, one can imagine other such contribution theories, or variations on those existing. Each theory will be described briefly in turn.

Goal-Contribution Theories

According to goal-contribution theories, the function of a part of a system consists in its contribution to a *goal* of that system. The notion of a "goal" or of a "goal-directed system" occupied a significant place in philosophical approaches to teleology from the 1940s through the early 1970s (Rosenblueth *et al*. [1943]; Sommerhoff [1950; 1969]; Braithwaite [1953]; Nagel [1953; 1961]; Beckner [1969]; Manier [1971]). However, it largely fell out of favor among philosophers of biology in the early 1970s, partly owing to the predominance of evolutionary considerations within that tradition and partly owing to internal conceptual shortcomings (Wimsatt [1972, 20-22]; Ruse [1973, 181-190]; Hull [1974, 109-111]). In short, a goal-directed system is one that exhibits a capacity to attain a specific value for some system variable, or to maintain the variable within a range of values, in the face of environmental perturbation, via the existence of compensatory activity operating amongst the system's parts.[93] The maintenance or attainment of a given value for the system variable is considered the *goal* of the system, and the specific contribution of a part of the system to that goal is considered to be the *function* of that part (Boorse [1976, 77]; Nagel [1977, 297]). Thus any system may have several goals;

[93] Such systems are often controlled by "negative feedback", but need not be; see below.

additionally, any sufficiently complex system can be analyzed as a hierarchy of goal-directed systems.

Two exemplary cases of "natural" purposiveness largely inspired this approach to teleology: homeostatic mechanisms drawn from physiology and servomechanisms that constitute the subject matter of cybernetics. As an example of the first type of mechanism, the percentage of water in the blood remains at around 90% throughout an individual's lifetime. This is because if it drops far below this level, the muscles increase the rate at which they infuse the blood with water; if it rises far above this level, the kidneys increase the rate at which they extract water from the blood. In this manner, the constancy of the water level of the blood is not a static phenomenon; it is actively maintained via compensatory mechanisms that operate throughout the body in the face of perturbation. Servomechanisms, such as heat-seeking missiles, exhibit a similar capacity to actively maintain a specific trajectory in the face of perturbation, and to adapt that trajectory to the moving position of the target. The oft-repeated slogan that goal-directed systems exhibit "plasticity" and "persistence" (e.g., Nagel [1977, 272]; Enc and Adams [1992, 650]) captures two central features of the concept of goal-directedness. On the one hand, such systems exhibit *plasticity* in that the same effect can be reached from a number of initial systemic configurations and by virtue of a number of different mechanisms or pathways. On the other hand, such systems *persist* in their course of action to the extent that they have the ability to attain or maintain a course of action in the face of environmental perturbation.[94]

[94] It is sometimes argued that the goal-supporting account does not allow one to determine a system *goal*, and consequently, that this goal must be arbitrarily stipulated (Wimsatt [1972, 20-22]; Schaffner [1993, 367-8]; Schlosser [1998, 327]). However, the above examples show this claim to be inaccurate. In the homeostatic case, *that* maintaining the water content of the blood at around 90% qualifies as a "goal" of the system is a consequence of the definition of "goal", in addition to a rudimentary understanding of physiology. It need not be arbitrarily stipulated.

The concept of goal-directedness has often been analyzed narrowly in terms of *negative feedback*, since such systems are capable of exhibiting self-regulation (Rosenbleuth *et al*. [1943]; Manier [1971]; Adams [1979]; Faber [1984]). In short, a "feedback" system is one such that a portion of that system's output is passed, or "fed back" into that system as input (but see Wimsatt [1971] for criticism of the concept of "feedback"). Feedback which tends to reduce or stabilize output is "negative"; feedback that increases output is "positive". However, theories of goal-directedness that emphasize the compensatory and self-regulatory activity of systems are not necessarily tied to negative feedback. Hull (1973) points out that a system can exhibit the plasticity required to be goal-directed without being guided by negative feedback. For example, if the kidney does not succeed in ridding the body of excess water, then sweating may do so, but the different responses are not obviously regulated by negative feedback, but by the utilization of alternative pathways (Ibid., 110-111). (Nagel [1953, 211], Sommerhoff [1969, 198-9], and Schlosser [1998, 309], also point out limitations of the negative feedback model for analyzing goal-directedness.)

Recently, Schlosser (1998) adopted some of the basic insights from the goal-supporting theory while rejecting its association with negative feedback (Ibid., 309) – although, strictly speaking, his theory should not be conflated with a goal-contribution view. According to his view, if a state or property of a system has a function then there exists a set of circumstances under which it is necessary for its own "re–production" – that is, its trans-generational reproduction or intra-generational persistence (Ibid., 326). However, in order to avoid the Boorse-type counterexamples described above, he stipulates that the system in question must be capable of *complex* self–re–production – that is, the system must be capable of re–producing the state in different ways, depending upon the environmental circumstances (Ibid., 312). Hence his view incorporates the

plasticity criterion associated with goal-supporting theories while leaving indeterminate the mechanisms by which this plasticity is realized. Section 4.1.2 will elaborate and criticize Schlosser's notion of "complexity".

Two main problems afflict goal-contribution theories, the "problem of vacuousness" and the "problem of goal-failure". The problem of vacuousness stems from the fact that the standard characterization of a goal-directed system as one that exhibits "plasticity and persistence" with respect to a given end is not sufficient for imposing a substantive distinction between different types of systems, for almost all systems can be described as seeking an equilibrium state which can be reached from different initial states and in different ways (Wimsatt [1971]; Woodfield [1976]; Nissen [1980-81]; Bedau [1993]). A pendulum swinging to a state of rest, a ball rolling from the top of a bowl to the bottom, and an elastic solid returning to its original condition after the imposition of tension would all represent goal-directed systems. Consequently, unless one specific mechanism, such as negative feedback, is included within the definition, it is difficult to exclude such counterexamples. Sommerhoff (1950, 86), and Nagel (1977, 273), attempt to exclude such systems by imposing an independence condition on the variables, which roughly states that all of the controlling variables must be independently manipulable; Schlosser (1998) attempts to do so by incorporating the "complexity" criterion stated above.

The problem of goal-failure stems from the fact that most explications of goal-directedness have tacitly or explicitly assumed that the supposed goal-directed behavior is successful, and as a consequence it is not clear how to explain the intuition that a non-conscious entity can have a goal and yet fail to satisfy it (Scheffler [1966 (1958)]; Beckner [1959]; Hull [1973]). Manier (1971, 234) and Adams (1979, 506) address this problem by arguing that what makes a negative feedback system "goal-directed" is not

that it actually achieves its goal, but that it is governed by an internal representation of the goal-state. This, however, raises the additional onus of providing a naturalistic account of "representation".

Good-Contribution Theories

The core idea behind the good-contribution view of functions is that in order for an entity to possess a function, performance of that function must (usually or typically) have a *beneficiary*. It must be useful for, beneficial for, or otherwise represent a "good" for some agent or system. This type of teleology is fairly evident in the world of artifacts, because artifacts are produced for a purpose and hence for an end deemed useful or beneficial by someone, and hence is closely associated with the mentalistic view described in Section 3.1.2.

However, this doctrine is not identical with the mentalistic view that all function ascriptions presuppose the existence of minds, because it is not incoherent to ascribe "interests" or "goods" to biological entities that cannot be said to possess the sort of mental life required by mentalism. For example, it is often said that natural selection preserves traits that are "beneficial" to their bearers. In fact, sometimes when a philosopher presents a theory of function according to which the function of an entity consists in its contribution to *fitness*, what is really intended is that the entity has a function only insofar as fitness is "good for" or "useful for" its bearer. Clearly, these theories are *evaluative* in the sense of Hare, as described in Section 1.3, in that any intelligible function ascription presupposes an act of commendation on the part of the person who utters it – i.e., that life is preferable to death, or that survival and reproduction are good things.

Canfield (1964), for example, defines the function of an entity simply as some useful contribution it makes to a system: "A function of *I* (in *S*) is to do *C means I* does *C* and that *C* is done is useful to *S*" (Ibid., 290). However, he argues, in the biological context usefulness can be translated in terms of a trait's making a contribution to the survival or reproductive capacity of its bearer (Ibid., 292). Sorabji (1964) also expounds a good-contribution theory, and he argues that Plato and Aristotle hold such a view. Ayala (1970) amends his etiological analysis by incorporating the concept of "utility" into his account: a feature of a system is "teleological" if it possesses "utility for the system in which it exists and such utility explains the presence of the feature in the system" (Ibid., 45). Thus, in the terminology adopted here, Ayala's position is, strictly speaking, a consequentialist one.

Presumably, one of the main advantages of such a view is that it appears to bridge the divide between natural functions and artifact functions, for, whereas artifact functions are "useful" by virtue of conscious design, natural functions are "useful" by virtue of their fitness contribution. In other words, the same concept is instantiated differently depending on the context, and hence there is no deep conceptual discrepancy between the usages. However, a significant problem with the good-contribution view is that it does not allow functions to be distinguished from "fortuitous benefits" or "lucky accidents". Frankfurt and Poole (1965), for example, criticize Canfield (1964) because heart sounds sometimes *do* have good consequences for fitness by alerting a physician to a potential life-threatening ailment, yet it does not have this as a function. (Wright [1973, 145-6] and Bedau [1992, 787] also raise this problem.) One solution to this would be to incorporate a statistical component: in order to have a function, the activity in question must usually, or typically, contribute to some good. But as Millikan (1984, 29) famously points out, statistical normalcy is not a reliable guide to functionality, since the probability that a

given sperm will actually fertilize an egg is extremely low, yet fertilization is without doubt the function of sperm. Most sperm are quite literally good for nothing. Finally, of course, accepting something like the good-contribution view would most likely spell the death of the project of "naturalizing teleology", since the ascription of function would be explicitly value-relative, and values are notoriously difficult to situate within the natural world.[95]

Bedau (1992, 794), like Ayala (1970), suggests the possibility of a theory of biological teleology that conjoins the etiological view and the good consequence view and that would ameliorate the problem of fortuitous benefits. According to this view, a trait would come to possess a function because its persistence is partly *explained by* its contribution to a beneficial consequence (increased fitness). However, he does not go so far as to offer an unqualified endorsement of this view, since the *goodness* of the result (increased fitness) does not itself perform an essential explanatory role in the etiology of the trait, but is only, as it were, externally linked to that explanation (Ibid., 801-2). McLaughlin (2001), however, develops a view according to which, in order to have a function, a trait must have produced a beneficial consequence that contributed to its own persistence or reproduction (Ibid., 168).

Fitness-Contribution Theories

The basic, unqualified idea behind fitness-contribution theories is that the function of a trait consists in its contribution to the fitness of the organism (or, more generally, to the fitness of the biological system of which it is a part). Thus, according to this view, the ascription of a function to a trait does not explain why that trait currently

[95] As such, it also seems to entail the peculiar consequence that if widespread attitudes about the value of survival were to change – that is, if humanity were placed in such a bleak situation that survival itself became a burden rather than a benefit – then function ascriptions would no longer be true!

exists, although ascription of a function to *ancestral* tokens of a trait can play a role in an explanation for the *current* persistence of that trait. Fitness-contribution views are proposed by Canfield (1964), Lehman (1965), Ruse (1971, 1973), Bechtel (1986), Bigelow and Pargetter (1987), Horan (1989), Walsh (1996), and Wouters (2003, 2005a, 2005b) (although, as pointed out above, Canfield [1964] accepts this view insofar as he defines function in terms of utility and believes that the fitness contribution made by a trait is "useful" to the organism). Sarkar (2005, 18) presents a generalization of this view, according to which a part of a system must merely contribute to the persistence of its containing system in order to have a function, and not necessarily to the reproduction of that system. This would allow functions to be assigned to the parts of, e.g., sterile organisms (see Section 3.1.2., *fn*. 74).

One problem with this unqualified view is that, in principle, fitness assignments can vary wildly depending upon fluctuations in the current environment, but function assignments tend to be relatively stable. For example, even traits that are, on average, adaptive in a given environment can, in certain environments, become maladaptive. But it is not said that in such an environment the trait no longer has a function, but that it is unable to perform its function.

Such counterexamples suggest that such function ascriptions should be relativized to a "normal" or "average" environment, in order to exclude abnormal or transient ones. This recognition led Bigelow and Pargetter (1987) to propose that a trait has a function when it bestows a survival-enhancing propensity on the organism that possesses it, in that organism's natural habitat (Ibid., 192). Thus, their definition of function introduces a counterfactual element – if the trait *were* in its natural habitat, then it would, *ceteris*

paribus, contribute to the fitness of its bearer. Yet this introduces further problems. Obviously, the "natural habitat" for an organism is not necessarily the organism's *current* habitat. But if not, then what constitutes an organism's natural habitat? One candidate for the natural habitat of an organism is that habitat in which it has, historically, flourished (Millikan [1989b, 300]; Mitchell [1993, 258-9]; Godfrey-Smith [1994, 352]; Walsh [1996, 562]). But then the propensity theory of functions is rendered perilously close to some version of the etiological theory, since its incorporation of a historical component violates the spirit of the "forward-looking" view they endorse.[96]

Walsh (1996; also see Walsh and Ariew [1996]) attempts to eliminate the problem of specifying the organism's natural habitat by proposing a "relational theory" of function, according to which this fitness contribution must be relativized to a specific "selective regime", which may have occurred in the past or the present. Hence, in his view, there are no functions *simpliciter*; in order to assign a function one must state precisely the nature of the environment within which the trait contributes to fitness. A problem with this view is that, as noted above, traits that are typically adaptive may become maladaptive in unusual environments, in which case it is often said that the trait is *unable to perform its function*. However, according to this view, one would have to say that, relative to the unusual environment, the trait simply does not have the function in question.

[96] However, this move would not, strictly speaking, make the view into an etiological one, since the historical dimension it introduces is not necessarily part of an explanation for the existence of anything. (Boorse [2002, 86; see *fn*. 26 of that paper] makes the point that introducing a historical dimension into the analysis of function does not entail introducing an etiological element; his view will be expounded in Section 3.2.1, below.)

A similar problem stems from the following consideration. In order to estimate the contribution of a trait to fitness, one must compare the average fitness of organisms that possess the trait with the average fitness of those that do not. But if no variation for that trait currently exists – such as the human kneecap – then it is not clear what to compare its performance with (Frankfurt and Poole [1965, 71-2]; Wimsatt [1972, 55-61]; Millikan [1989a]; Godfrey Smith [1994, 352]). One possibility would be to compare it with the variation that existed at an *earlier* time. But again, this brings the propensity theory closer in spirit to the etiological view.

Wouters (2003, 2005a, 2005b) proposes a version of the fitness-contribution view according to which, in order to have a function, a trait must confer a biological advantage upon its possessor, relative to some actual or counterfactual set of variants. This resolves the problem insofar as one must explicitly stipulate the range of variation in question. Moreover, he argues that this reflects standard practice within some fields of biology. In optimality models of adaptation, for example, the relevant range of alternatives (the "phenotype set") is typically derived from biologically-informed assumptions about what is physically, ecologically, or physiologically possible (Parker and Maynard Smith [1990, 27]; also see Wouters [2005a, 43]). However, merely stipulating the range of variants in question seems to introduce an element of arbitrariness into function ascriptions. Relative to one hypothetical set of variants, a trait has a function; relative to another set, it does not. Clearly, something more substantive should be said about how this range of variation can be determined in a biologically plausible manner.

As noted above, the main advantage of contribution-based theories is that they are more consistent with the majority of biological usage. Moreover, given the difficulty of

inferring the evolutionary history of a trait from its current activity, it makes the practice of ascribing functions much more amenable to empirical testing. However, these theories appear to deprive functions of two of the properties that have, historically, been associated with their use and that continue to be associated with them. The first is that they are explanatory in the sense that they specify an efficient cause for the current existence of the trait. What this means is not that the fact that a trait *had a function* in the past explains its current existence, but a trait's *having a function* explains its current existence. The second is that they are normative. On the etiological view, the distinction between functioning properly, malfunctioning, and inability to function due to an abnormal environment, is rendered tolerably clear: because of the fact that function is a historical concept, something can have a function without being able to perform it. It is controversial whether these distinctions can be drawn clearly within consequentialist theories; though it has been argued that consequentialist views can sustain normative interpretations of function (Wimsatt [1972, 47]; Walsh [1996, 568]; Schlosser [1998, 327]).

These considerations reinforce the value of adopting a pluralistic and context-dependent approach to analyzing "function". In other words, it is not so important to define what "function", as such, *means*. It is more important to be able to evaluate, for any given scientific usage, what inferences that particular usage is intended to support or is capable of supporting, and then to specify which concept of function most adequately permits those inferences to be drawn. This dissertation argues that in biological psychology, any theory of function must be normative in a non-externalist manner, and

that only etiological theories satisfy those constraints. This will be shown in the next section. In other disciplines, these conditions need not be so restrictive.

3.2 UNIQUENESS CLAIM FOR ETIOLOGICAL THEORIES

The purpose of this section is to argue that etiological theories of function are uniquely capable of satisfying both adequacy conditions CA_1 and CA_2*.[97] What this means is that consequentialist accounts of function violate one or the other conditions, and that the etiological theory is consistent with both.[98] The first part of the section (Section 3.2.1) shows that consequentialist theories – or at least those that have been presented in the literature – cannot satisfy both conditions. This does not mean that it is *impossible* for a consequentialist theory to do so, but that two well-developed attempts to define "dysfunction" from a consequentialist perspective fail. The second part of the section (Section 3.2.2) will show that etiological theories can. It will do this by using the etiological theory to construct a definition of "dysfunction", and then showing that this definition satisfies both conditions.

3.2.1 Consequentialist Theories Violate Adequacy Conditions

This subsection will show that consequentialist theories of function are incapable of satisfying both adequacy conditions. The argument will not proceed, theory by theory, to show that each theory violates one or the other condition. Rather, it will use the distinction drawn above between methodological and substantive approaches to show that

[97] Although representationalist etiological theories are also capable of satisfying both adequacy conditions, that will not be established here. The argument that will be presented only applies to non-representationalist forms, because it relies crucially on a stipulative definition of a "normal environment for a trait's functioning" that does not necessarily apply to artifacts.

[98] Clearly, if a theory does not satisfy CA_1 then it cannot satisfy CA_2* either.

whichever one is selected, a concept of function that satisfies both adequacy conditions cannot be defined within that approach.

Methodological Approaches Violate Adequacy Conditions

As noted above, interest-based contribution theories are methodological. In the following, Cummins' (1975) analysis will be taken as representative of this approach and criticized. It will be argued that it cannot satisfy CA_2*. However, the criticism will clearly apply to all methodological approaches *as such*.

On the face of it, the functions bestowed by Cummins-style functional analysis cannot satisfy CA_1 for the following reason. If an entity is dysfunctional then it is not performing one of its functions. This implies that it does not perform some activity that plays a salient role in the context of an analysis of some system capacity. But it is capable, of course, of performing *some* activity, and this activity could, in principle, play a salient role in the context of an analysis of some *other* system capacity – other than the one that is currently the subject of interest. In Cummins' account, relative to the capacity of an individual to experience excruciating pain, the function of the kidney is to support calcium formations along its inner wall. Whether or not harboring kidney stones is the function of the kidney, then, depends upon one's explanatory interest. This suggests that, though a concept of dysfunction *can* be defined for this methodological approach, it is a relative, and not an absolute, concept. (That Cummins-functions are capable of defining "malfunction" in this interest-relative way was noted by Godfrey-Smith [1993, 200], and Hardcastle [2002, 152]).

The problem with methodological approaches to function, then, is not that they cannot satisfy CA_1, but that they cannot satisfy CA_2*, since *being picked out as a salient feature within the context of a functional analysis of a system capacity* qualifies as an externalist criterion. This is because, whether or not an activity of a trait is so picked out

depends upon epistemological considerations that do not necessarily have any effect upon the characteristic structure or activity of the trait itself. The peripheral capillaries contribute to both the circulatory system and to the thermoregulatory system, and will consequently be ascribed different "functions" depending on which system is selected; nonetheless, the capillaries continue to dilate and contract, regardless of whether a scientist prefers to analyze its activity in relation to the circulatory system or the thermoregulatory system. Hence, changes in the focus of scientific interest qualify as *external* to the system in which that interest is taken. This suggests that if one is to construct an adequate concept of "dysfunction", one will have to appeal to substantive, and not methodological, concepts of function.

Substantive Approaches Violate Adequacy Conditions

Goal-contribution, good-contribution, and fitness-contribution theories represent substantive approaches. Other contribution theories have been mentioned – persistence-contribution theories, for example. Presumably, one could find other plausible contribution theories. Consequently, owing to the open-ended plurality of such theories, the strategy that will be pursued here will *not* consist in enumerating the problems that each type of substantive contribution theory faces when it comes to defining "dysfunction". Rather, it will be to criticize two very general, theory-independent methods that advocates of substantive contribution theories have used to explain the possibility of dysfunction.

The general problem that consequentialist theories face is this: in order to *have* a function, a part must contribute in the appropriate way to the specified outcome. But by definition, if a part is dysfunctional it cannot so contribute. Therefore, the function of a

part cannot simply consist in its contribution to the specified outcome.[99] There are two ways in which consequentialist theories have attempted to avoid this problem. The first is to define the function of an entity in terms of a statistical measure taken over the activities or capacities of a class of entities of the same type. The second is to define the function of an entity in terms of its *disposition* to perform the activity in question. The statistical account will be given first, then the dispositional account. Both will be shown to violate CA_2*, since, in order to evaluate whether a given token of a trait is dysfunctional or non-dysfunctional, the statistical account essentially relies upon what *other* tokens are doing, and the dispositional account is forced to identify the distinction between a dysfunctional and non-dysfunctional trait with that between being adaptive and being maladaptive, which may admit of purely externalist determination.

Statistical Account of Function

A statistical account of function is one that defines the function of a given token of a type in terms of a statistical measure taken over entities of that same type. An example of such an account, suggested above (see Section 3.1.3, under "Good-Contribution Theories"), is the identification of the function of a trait with that activity that *frequently* contributes to the good of a containing system. Clearly, such an approach can satisfy CA_1, for, since it assigns a function to an entity on the basis of what other entities do (in addition, perhaps, to its own activity), it allows an entity to have a function without being able to perform it. Wimsatt (1972), Boorse (1975; 1976; 1977; 2002), Walsh (1996), and Walsh and Ariew (1996) all appeal to a statistical norm in order to explain the possibility of non-performance of a function.

[99] Certain philosophers who hold a consequentialist theory, such as Davies (2000), Boorse (2002, 89), and Wouters (2005b, 128), simply deny that an entity can have a function that it cannot perform and hence they reject the concept of "dysfunction" altogether (see *fn*. 100).

Boorse, for example, includes a statistical component within his goal-contribution account of function, according to which the function of an entity consists in its species-typical contribution to a goal. In physiology, this "goal" is identified with survival and reproduction; hence, "the specifically physiological functions of any component are…its species-typical contributions to the apical goals of survival and reproduction" (1975, 57; also see 1976, 77; 1977, 556). Hence, a token can make a one-time, accidental contribution to survival without that contribution constituting its function (1977, 556-557). Moreover, this definition also allows for an entity to have a function without being able to perform it.[100] He also extends the notion of a "species-typical contribution to survival and reproduction" to include not only what is typical for the present moment, but what has been typical over a longer time period that also includes the past. This explains the possibility of pandemic diseases, that is, diseases that bring about functional impairments in the majority of present tokens of a trait (Boorse [2002, 99]). For example, one explanation for the rapid decline of the Cascades frog, *Rana cascadae*, and the Boreal toad, *Bufo boreas boreas*, in Oregon appeals to the recent increase in ultraviolet B radiation, which appears to affect immune response and hence render larvae more susceptible to a regional fungus (see Sarkar [1996] on recent anuran declines). Even though this susceptibility may appear "normal" if one considers only the present time (if the majority of such anurans are affected) it is highly unusual or abnormal if one considers a longer time scale that extends into the past.

Similar to Boorse, Walsh (1996; also see Walsh and Ariew [1996]) defines the function of an entity, relative to an environment, in terms of the (positive and significant) contribution it makes to the *average* fitness of individuals with that trait: "The/a function

[100] Boorse, however, does not offer a definition of "dysfunction"; in fact, he rejects the idea that something can possess a function that it is constitutionally unable to perform (2002, 89). Instead, he uses his definition of "function" to define "disease" as the reduction of a functional ability below typical efficiency (1977, 562). Hence, this dissertation puts his concept of function to a use that goes against his express intent.

of a token of type *X* with respect to selective regime *R* is to *m* iff *X*'s doing *m* positively (and significantly) contributes to the average fitness of individuals possessing *X* with respect to *R*" (Walsh [1996, 564]). (By "selective regime" he refers to all environmental factors that potentially affect an individual's fitness [Ibid.]). On this basis, he explicitly defines "malfunction" in terms of the inability to make this contribution: "Those individuals whose *x*'s can't do *m* suffer a fitness decrement, on average. Their tokens of *X* malfunction (no matter what else they are capable of doing)" (Ibid., 568).[101]

Note that both of these statistical approaches resolve the problem mentioned above (Section 3.1.3, under "Good-Contribution Theories") that, from a statistical point of view, a given individual that has a function may have a very low probability of actually performing that function – such as the probability that a given sperm will fertilize an ovum. The reason that it avoids this problem is that it does not define the function of a trait in terms of what that trait frequently *does*, but in terms of what it frequently does *when* its activity contributes to fitness (Boorse [2002, 93]). In other words, to evaluate the function, *F*, of a trait, *T*, one must calculate the number of times in which *T*'s doing *F* contributes to fitness over the number of times in which *T*'s doing *something* contributes to fitness. If one restricts one's attention to those individual sperm

[101] The purpose of Walsh's relativization to a selective regime is to allow his theory to remain neutral between "backwards-looking" and "forward-looking" accounts of function: if the relevant environment is a *past* one, then function statements help to explain the current maintenance of a trait in a population; if the relevant environment is the present one, then the account is a consequentialist one (Ibid., 570). He believes that, like the etiological theory, the relational theory allows function statements to be explanatory (Ibid., 571) as well as normative (Ibid., 586), but that since it is not inherently historical it is more consistent with biological usage (Ibid., 558). However, as noted above (see Section 3.1.3, under "Fitness-Contribution Theories"), it is important to make a distinction between the claim that a function that a trait had *in the past* helps explain its current maintenance, and the claim that a trait's *presently* having a function helps to explain its current maintenance. Unlike the etiological view, the relational view cannot account for this latter fact; hence, according to the relational view, the present-tense ascription of a function does not explain anything. Similarly, the purpose of this section is to show that, to the extent that it permits the concept of dysfunction to be defined, then the relational view is *normative*, but it is also externalistic in that it violates CA_2*.

that *have* contributed to fitness, the vast majority have done so by fertilizing ova, and hence that constitutes its function.

Of course, this solution allows for vagueness: how high must this proportion *be*? On the one hand, the function of a trait cannot simply be defined as whichever one of a trait's activities *most frequently* contributes to fitness, since this does not allow a trait to have more than one function unless it performs each of its functions with equal frequency. The medulla oblongata regulates breathing by monitoring CO_2 and stimulating the diaphragm and intercostal muscles; it also induces vomiting. Both are performed over the course of a normal life and both are essential to survival, yet the medulla contributes to fitness much more frequently by regulating breathing than by initiating vomiting. Nonetheless, both contributions are equally considered its "functions". On the other hand, it seems arbitrary to specify a given probability, *p*, and stipulate that the trait must perform the activity more frequently than *p*. This is especially so because some traits that typically contribute to fitness by performing an activity *m* may also have the function of performing another activity *m** in rare or exceptional circumstances – as noted above, the frequency with which the medulla contributes to fitness by initiating vomiting is extremely low compared to the total number of occasions in which it contributes to fitness, so it does not seem that from a biological point of view there are any principled reasons for imposing an absolute limit on this proportion. But this problem of vagueness is not restricted to fitness-contribution theories. It is faced by etiological theories as well – for example, how often must a trait have contributed in the past to the relative fitness of individuals possessing it before it is "selected for"?[102] There does not seem to be a non-arbitrary answer to this question.

[102] Boorse (2002, 71) points out that, *contra* Millikan (1993, 35-39), to say that a trait was "selected for" does not eliminate the vagueness in specifying the number of past generations over which its activity must have conferred a relative fitness advantage upon its bearer in order for it to have a function. For example, if there was only a single occasion in which a trait variant allowed its bearer to reproduce more effectively

In the present context, the statistical approach faces a much more important problem than vagueness. Because it defines the concept of "dysfunction" in terms of the species-typical contribution of a trait, it admits of externalistic ascription – that is, the difference between a trait's being dysfunctional and being non-dysfunctional does not necessarily supervene on the characteristic structure and activity of that trait. In order to show this, one must construct a scenario in which a given token, *t*, of a trait type, *T*, is unable to perform the activity, *F*, that constitutes *T*'s species-typical contribution to fitness – rather, *t* does *F** instead, which contributes to fitness very rarely. In this environment, *t* is "dysfunctional". Then, one changes the environment in such a way that *F* is no longer associated with a fitness advantage, but the fitness of the organism possessing *t* is unaffected, as is *t*'s characteristic structure and activity. By doing this, one increases the number of cases in which *T* contributes to fitness by doing *F** over the total number of cases in which *T* contributes to fitness, and hence *T* loses the function *F*, and gains the function *F**. As a consequence *t* is no longer dysfunctional, but its characteristic structure and activity have remained unaffected by this transition. The trait variant has gone from being dysfunctional to non-dysfunctional simply by altering the fitness of a different variant. A very similar problem was encountered in the context of the previous discussion of Kendell's (1975b) concept of disease as "biological disadvantage" (Section 2.3.1). Consequently, statistical approaches to function are not capable of satisfying CA_2*.

Dispositional Account of Function

---

than the bearer of another, then one would not say that it came to possess a "function". So how many generations are required before a trait has been "selected for"?

Bigelow and Pargetter (1987) present a dispositional approach to function. Even though they advocate a variant of the fitness-contribution theory, the criticism that follows can be extended to the other substantive contribution theories as well. According to their definition of function, "something has a (biological) function just when it confers a survival-enhancing propensity on a creature that possesses it" (Ibid., 192). More precisely, they define "function" counterfactually, and relative to the natural habitat of the organism, for reasons explained above (Section 3.1.2, under "Fitness-Contribution Theories): a function "*would* give a survival-enhancing propensity to a creature in an appropriate manner, in the creature's natural habitat" (Ibid., 193). A "propensity" is a type of disposition that is manifested probabilistically, and will be described in more detail below.

Note that the disposition to survive is not a property of the functional entity itself, but of the organism: *given* the way the entity is structured, if the organism *were* in its natural environment, the entity would bestow a disposition to survive upon the *organism*. A short explanation of the concept of a disposition (and that of a propensity) is warranted for the sake of criticizing their position.

A paradigmatic example of a dispositional property is the solubility of salt. To say that salt is soluble is to say, among other things, that *if* salt is placed in water, then (all things being equal) it will dissolve. Moreover, the reason salt is soluble is that it has a polar molecular structure. When immersed, the hydrogen end of the water molecule attaches to the chloride atom, and the oxygen end the sodium atom. This breaks the sodium chloride bond. Dispositions are not *de novo* properties of complex systems; rather, they supervene upon the structure of those systems (Prior *et al*. [1982]).

This example illustrates two properties of dispositions. The first is that they can only be defined relative to a given set of circumstances. Often, in addition to the

circumstances that are explicitly specified in the definition (water-immersion) there is a *ceteris paribus* clause that places additional, implicit restrictions on those circumstances. For example, salt will not dissolve if the water is saturated, or if a non-soluble container protects it, and so on. For the present purposes it is irrelevant how many such conditions must be included in the set of circumstances, or whether the number of such conditions is even finite.

Note, however, that dispositions are not relative in the sense that whether or not an entity *has* the disposition depends upon whether it is presently within the circumstances in question. Salt is soluble even if it is never placed in water. This is crucial for understanding why Bigelow and Pargetter define functions as dispositions. It allows something to *have* a function even if it cannot perform it, in the same way that salt has a disposition even if it cannot manifest it (Bigelow and Pargetter [1987, 193]). Hence, their definition can satisfy CA_1.

The second property is that something has the disposition it does because it has the structure it does. The solubility of salt, as noted above, is a consequence of its molecular structure: it supervenes on this structure. A detailed knowledge of the structure that explains the disposition clearly enhances the explanatory power of dispositions. In fact, when *no* knowledge of the structure of the entity with a disposition is possessed, explanations that appeal to dispositions are often suspected of being vacuous. The classic example of a vacuous explanation is taken from Molière's seventeenth-century parody, *Le Malade Imaginaire*. When asked why opium puts people to sleep, the learned doctor explains that opium has a *virtus dormitiva*, the Latin phrase for "sleep-causing power". This is not unlike explaining why the glass broke by saying that it is fragile.

Regardless, it is irrelevant for the present purposes whether appealing to a disposition without structural knowledge is or is not explanatorily vacuous. What matters

in the present context is that, to ascribe a disposition, *D*, to an entity of type *X* is to imply that for all entities, *X*, that have *D*, there exists a structure, *S*, such that *X* has *S* and, because *X* has *S*, if *X* is placed in environment *E*, then (*ceteris paribus*) it will manifest *D*.[103] (This is not intended as an *analysis* of "disposition".)

A "propensity" is a probabilistic, or variable-strength, disposition. Presumably, propensities are correlated with frequencies, in the sense that, if things of type *X* have a higher propensity than things of type *Y* to manifest a disposition, *D*, under conditions *C*, then *X*s will manifest *D* under *C* more frequently than *Y*s will manifest *D* under *C*. One might question, however, whether or not "propensity" is a very useful concept. This is because whether or not an object manifests a disposition is usually thought to be *completely* determined by its structure and its environment. *If* salt is placed in water, *and* all of the implicit *ceteris paribus* conditions are met, then it will dissolve. It will not "probably" dissolve if these conditions are met; it will definitely dissolve. To say that salt will probably dissolve could only mean that one does not know if all of the relevant circumstances actually hold. However, propensities *are* useful for describing the dispositions of genuinely indeterministic phenomena, such as the decay time of some sub-atomic particles. In principle, two identical radioactive atoms placed in two identical environments can decay at different times.[104] If this phenomenon is inherently indeterministic then there does not exist an exhaustive description of its structure and environment from which the time of decay can be determined; hence, one can only attribute a propensity to it.

---

[103] Note that this condition is consistent with the possibility that dispositions are multiply realized; that is, that there is no unique structure that must be associated with a given disposition.

[104] According to an influential definition of "fitness" (Mills and Beatty [1979]) the fitness of an individual *organism* consists in its propensity to leave a given number of offspring (Ibid., 275). They interpret this propensity as its expected number of descendents in an environment (hence it is not a probability but an expectation). This interpretation should not be taken to imply that the number of offspring an organism leaves is genuinely indeterministic. Consequently, their definition could, in principle, be redefined in terms of dispositions rather than propensities.

On Bigelow and Pargetter's account, as noted above, the disposition is not ascribed to the functional entity itself, but to the organism: a trait has a function if it bestows upon the *organism* a disposition to survive in that organism's natural habitat. Presumably, this implies that a trait, *T*, has a function when the organism has a structure, *S*, (which includes possession of *T*, along with other species-normal parts and processes) such that, if the organism is placed in environment *E*, namely, its natural habitat, then (*ceteris paribus*) it will survive. Contrary to appearance, this statement does *not* imply the strong claim that the organism with the trait in question will *always* survive in its natural habitat. In order to define a concept of "dysfunction", then one must ask: on the dispositional view, under what conditions will a functional trait fail to contribute to survival? There are four possible conditions:

> (i) the organism is in *E* and possesses *S*, but some condition left implicit in the *ceteris paribus* clause is not satisfied – for example, a new predator is introduced into the habitat. However, for the purpose of defining "dysfunction", this possibility may be ignored, since, in principle, one can simply interpret *E* broadly to include the *ceteris paribus* condition;[105]
>
> (ii) the organism is in *E* (construed broadly to include the *ceteris paribus* condition) and possesses *S*, but since a propensity, unlike a disposition, is

[105] Of course, one might not, in a given case, *know* what *ceteris paribus* condition went unfulfilled, and hence one may not be able to render it explicit. The point, however, is that whether or not an entity is dysfunctional should not be determined by whether or not the organism failed to survive in its natural habitat because an unknown *ceteris paribus* condition was violated. The reason is that whether or not a given *ceteris paribus* condition is *known* is an *epistemological* concern – that is, it is relative to the knowledge that is available to one – and hence if it were used to define the difference between a dysfunctional and non-dysfunctional token of a trait, the definition of "dysfunction" would be an externalist one, since the characteristic structure and activity of a trait is not necessarily affected by how much information a person has about it.

probabilistic, the organism may still fail to survive. This option entails that whether or not an organism survives in its natural habitat is genuinely indeterministic. But imputing the sort of indeterminacy found in the quantum realm to the organismic realm is physically implausible. It would entail that, given two identical organisms in two identical environments, one might live and the other die. Although it is possible, such a scenario is so unlikely that a concept of "dysfunction" defined on its basis would be vacuous;

(iii) the organism does not have *S*. But if the organism does not have the requisite structure, then it does not possess the disposition in question, since disposition supervenes on structure. Consequently, the trait in question is not able to bestow a survival-enhancing disposition upon the organism, and hence, by definition, it does not have a function;

(iv) the organism is not in *E* (construed broadly to include the *ceteris paribus* condition). According to this option, even if the structure *S does* bestow a survival-enhancing disposition upon the organism, the current environment of the organism is not such as to allow the disposition to be manifested. A glass is fragile even if never struck; an organism is disposed to survive even if it is not in an environment conducive to the manifestation of that disposition. This last alternative will be used in the following to define a concept of "dysfunction", which will be shown to violate CA_2*.

One could define a concept of dysfunction on the basis of (iv) in the following way: a trait is dysfunctional if it bestows a survival-enhancing disposition upon the

organism, but the current environment of the organism is not such as to allow the disposition to be manifested. The problem with this formulation is that it violates CA_2*, since it permits the distinction between dysfunctional and non-dysfunctional token of a trait to be determined by factors that do not affect the characteristic structure or activity of the trait itself. An example will illustrate this.

One of the optic fibers of the frog, *Rana pipiens*, is differentially sensitive to the presence of flies in its environment (Lettvin *et al*. [1959]). In fact, any appropriately positioned small dark moving object will cause the frog to strike in the direction of that object. Suppose that there are two frogs, one of which is placed in its natural environment and the other in an artificial environment filled with small moving black objects that are all coated with a lethal poison. In both habitats, the characteristic structure and activity of the optic nerve are exactly the same in each frog: the optic nerve carries information about small dark moving objects to the superior colliculi, which guides the motor response to the position and direction of the stimulus. However, the latter environment is not one in which the survival-enhancing disposition it bestows upon the organism can be manifested; consequently, according to the definition of "dysfunction", the optic nerve of the first frog is functional and that of the second is dysfunctional. This violates CA_2*.

One might object that this example does *not* violate CA_2*, since the differences in the frogs' environments will *eventually* have an effect on the characteristic structure and activity of their respective optic nerves: eventually, the second frog will consume a poisoned object and die, and the first frog will consume a nutrient-rich fly and live. But the fact that the eventual outcomes of the two scenarios will diverge is not relevant for CA_2*, since the dispositional account of dysfunction allows one to say that the optic nerve of the second frog is dysfunctional even *before* any relevant physical changes take place that would distinguish them. The optic nerve of the second frog becomes

dysfunctional *as soon as* that frog is placed within the artificial environment, because at that time the frog is no longer in an environment in which the survival-enhancing disposition can be manifested. What this suggests is that the dispositional account of dysfunction is more closely related to biological usage of the term "maladaptive", which often simply means that, in a specified environment, a trait is fitness-reducing. But as noted in Section 2.2.2, this is an externalist concept.

Note, finally, that the dispositional account of dysfunction has a very counterintuitive consequence, although one that does not technically violate either of the adequacy conditions: it does not allow one to say that an entity is malfunctioning in the case where a structural anomaly prevents it from performing its function. Suppose, as in the famous experiments of Sperry (e.g., Sperry [1944]) the optic fibers of a toad are severed, the toad's eyes are rotated 180 degrees, and the optic fibers are allowed to regenerate. As is well-known, the initial point-to-point mapping of connections is restored after regeneration, such that the toad's visual field is inverted both horizontally and vertically. As a consequence, the toad always flicks its tongue in the direction opposite that of the lure. But in this case, the structure of the optic nerve is simply no longer such as to bestow a survival-enhancing disposition upon the organism: "When the lure was held above the head and a little caudad to the eye the animals struck downward in front of them and got a mouthful of mud and moss" (Ibid., 63). Therefore, it is neither functional nor dysfunctional, since it cannot be said to have a function at all.

Similarly, if acute heart failure in an individual is brought about by stenosis of the mitral valve (formed by irregular hardening of the flaps of the mitral valve, which restricts the flow of blood from the left atrium to the left ventricle) the heart cannot be said to "malfunction", since it no longer bestows a survival-enhancing disposition upon the organism, or at least not to the same degree as in its absence. Neander (1991a, 183)

makes this point in arguing that Bigelow and Pargetter's theory cannot define "malfunction" since, in paradigmatic cases of malfunctioning, the trait in question loses its ability to confer a survival-enhancing disposition upon the organism and hence does not have a function. As has been argued here, their theory *can* define "malfunction", but only for the case in which the environment does not permit the manifestation of the disposition, rather than that in which the structure changes in such a way that it cannot perform the activity. The notion of a "malfunction due to structural anomaly", according to the dispositional approach, is a contradiction in terms.[106]

3.2.2 Etiological Theories Can Satisfy Both Adequacy Conditions

The purpose of this section is to argue that non-representational etiological theories are capable of satisfying both adequacy conditions. In order to show this, it will use the etiological theory as the basis for a definition of "dysfunction", and it will show that this definition implies the truth of both adequacy conditions. The remainder of the dissertation will be concerned with extending the definition, resolving some conceptual problems with it, and showing how it applies to cases of interest in psychiatry.

According to the definition that will be presented here, to say that an individual entity, *x*, is dysfunctional with respect to function *F*, means:

(i) the function of *x* is *F*;

(ii) *x* is not able to perform *F*; and

[106] Their definition, then, is contrary to the one that will be presented in the following section (Section 3.2.2), in which an entity can *only* be dysfunctional when its structure is such as to prevent it from carrying out its function, rather than when the environment is not a permissive one. In the definition that will be presented here, "dysfunction due to an abnormal environment" is a contradiction in terms. See Section 5.1, where the distinction between being functional, dysfunctional, and unable to function due to an abnormal environment, will be elaborated and illustrated.

(iii) if *x* is not in the normal environment for its functioning, then if *x* were in the normal environment for its functioning, *x* would not be able to perform *F*.

Like Bigelow and Pargetter's (1987) definition of "function", there is a counterfactual element; that is, there is a reference to what would be the case were the entity in its normal environment. This, however, raises the onus of defining "normal environment for an entity's functioning", which was not attempted in the previous section. The etiological definition of a normal environment for an entity's functioning will allow the satisfaction of CA_2*, which has remained elusive so far.

That the etiological theory of functions satisfies CA_1 is easy to show, as indicated in the previous chapter, in relation to Klein's (1978) definition of "function" (Section 2.3.3). Since the etiological theory defines the function of an entity in terms of its history, then whether or not something *has* a function is independent of whether it is currently capable of *performing* its function, and hence it allows something to have a function without being able to perform it. Another way of putting this point is that *having* a function supervenes on history alone; *performing* a function supervenes only on the current structure and environment of the functional entity.

The argument for why etiological theories satisfy CA_2* is slightly more complex, but can be briefly summarized before it is elaborated in more detail. It rests upon two premises: a definition of a "normal environment for an entity's functioning" and a determination claim. First, the concept of a *normal environment for an entity's functioning* is defined as that environment within which past instances of the trait performed the activity that thereby came to constitute its function. This is the approach taken by Millikan (1984, 33-4). Second, it is assumed that the activity of an entity is determined by its *structure* and its *environment*. These two premises entail that *if*

something is unable to perform its function, then *either* its structure diverges from that of its progenitors, *or* its environment diverges (or both). If the *environment* diverges, then, by definition, the trait is not in its normal environment; but if it is the case that, were it in its normal environment, it would still not be able to perform its function, then its *structure* must diverge from that of its progenitors (since, by definition, the environment is the same). But then the only way of changing the functional status of an entity from dysfunctional to non-dysfunctional is to change its structure, and this is an internal change. Therefore it satisfies CA_2*. This argument will be elaborated in more detail below.

The first premise involves the definition of "normal environment for an entity's functioning". The definition of "dysfunction" given above specifies that if an entity is not in the normal environment for its functioning at the time of the ascription then it must be the case that the entity would still be incapable of performing its function even if it were in that environment (condition [iii]). The concept of "normal environment for an entity's functioning" is here defined simply as any one out of the range of environments within which the entity's progenitors performed the activity that currently constitutes its function, and in which those performances were fitness enhancing. The reason that the Arctic Circle, rather than the San Diego Zoo, constitutes the "normal environment" within which the functioning of the polar bear's dense, water-repellent fur can be assessed is because that is the environment in which, in the past, those properties contributed to heat retention and thereby explain its current existence.

Consequently, the notion of a "normal environment for an entity's functioning" is a retrospective, historical one.[107] This approach to defining the concept of normal

[107] This is *not* to imply that the concept of "normal environment for a trait's functioning" as defined here accurately reflects most biological usage. All that is necessary is that there exists a well-defined *reference environment* in relation to which functional activity of an entity – whether it is dysfunctional or non-dysfunctional – can be assessed.

environment was proposed by Millikan (1984, 33-34). She first defines the concept of a "Normal explanation" for the functioning of a trait. This is an explanation of how the trait has typically, historically, performed the activity that currently constitutes its function. It specifies the relevant structural features of the trait, and the relevant environmental conditions. The environmental conditions that are appealed to in a Normal explanation constitute "Normal conditions" for a trait's functioning.

Clearly, as she points out, Normal explanations can be more or less "proximate". One can describe a "Normal condition" for the heart's functioning as one in which oxygen is present in the organism's atmosphere. One can also describe this environment much more specifically as one in which the lungs deliver oxygen to the heart via the pulmonary vein. The latter explanation is more "proximate" than the former since it explains *how* the oxygen in the atmosphere comes to have an effect upon the heart.

Although the conditions under which a trait can dysfunction are only well-defined by constructing a well-defined normal environment, it is important to note that the normal environment for a *trait* is not necessarily the same as the normal environment for the *trait-bearer* (e.g., the organism). (Walsh [1996, 563] draws attention to this ambiguity in Bigelow and Pargetter's [1987] usage.) The normal environment for a functional *trait* consists in the (more) proximate physical environment that provides that trait with the materials that, in the past, contributed to that functional ability. For example, a normal environment for the human *heart* is not the normal environment for the *person*, but rather, the environment that consists in a supply of blood from the vena cava, the modulation of the heart rate by norepinephrine and acetylcholine from the sympathetic and vagal nerves, respectively; a supply of oxygenated blood from the pulmonary vein; and so on. If the lungs are not supplying oxygen to the heart, then the heart is *not* in its normal environment, even if the organism that possesses the heart can be said to occupy

its *own* normal environment – that is, even if there is oxygen present in the organism's atmosphere. By the same token, if the lungs are supplying oxygen to the heart, then (all things being equal) the heart *is* in its normal environment, even if there is no oxygen in the atmosphere. For example, the heart of astronaut receiving an artificial oxygen supply is in the normal environment for its functioning. Clearly, this notion of a more or less "proximate" normal environment for the heart's functioning needs to be rendered more precisely, but this suffices for the present.

With the concept of the "normal environment for an entity's functioning" in place, one can now show how the proposed definition of "dysfunction" satisfies CA_2*. It will be assumed that the *activity*, *A*, of a functional entity, *x*, is determined by its structure, *S*, and its environment, *E*. What this means is that *S* and *E* determine *A*. Now, suppose that an entity of the same type as its progenitors cannot perform the activity that its progenitors performed and that thereby constitutes its function. By the determination claim, this implies that one of the following three cases holds:

(1) *x*'s structure is different from that of any of its progenitors that performed *A*, but its environment is the same as one of its progenitors ($\neg$S & E). But if *x*'s *structure* is different and the environment is the same, then *x* is unable to perform its function in its normal environment and is therefore dysfunctional – it satisfies all three conditions;

(2) *x*'s structure is the same as that of one of its progenitors, but its environment is different from that of that progenitor (S & $\neg$E). But if *only x*'s environment is different, then *x* is in an abnormal environment; since its structure is the same, then *were x* to be placed in its normal environment (that environment within

which its structurally similar progenitor performed *its* function) it would be able to perform the activity in question. Consequently, merely being placed in an abnormal environment cannot satisfy (iii) of the definition of "dysfunction";

(3) *x*'s structure and environment are different from the structure and environment of any of its progenitors (¬S & ¬E). But if both *x*'s structure and *x*'s environment are different, then *x* is in an abnormal environment and one would have to assess whether it would still be able to perform its function in the normal environment. If its structure varies such that it could not perform its function, even if it were to be placed in its normal environment, then it is dysfunctional.

The following table (Table 3.3) illustrates the four possible circumstances for a trait's functioning. The most important point about Table 3.3 is that *the only context in which an entity can be dysfunctional is that in which its structure diverges from that of ancestral tokens that performed that activity*. Whether or not it inhabits its normal environment is irrelevant for the evaluation of its functional status. Consequently, the only way to change an entity from being dysfunctional to being non-dysfunctional is by changing its structure and not merely its environment. But structural changes are changes that are internal to the entity itself. Therefore this definition satisfies CA_2*.

	Structure similar to progenitor (S)	Structure not similar (¬S)
Environment similar to progenitor (E)	Necessarily non-dysfunctional	Possibly dysfunctional
Environment not similar (¬E)	Necessarily non-dysfunctional	Possibly dysfunctional

Table 3.3 Cases under which an entity can be dysfunctional. (See accompanying text for details.)

The argument given above is fairly simplistic. One obvious objection is that, strictly speaking, at any moment, the current environment of a trait, and the current structure of that trait, are different from that of every other moment. What are the identity conditions for "sameness of structure" and "sameness of environment" over time? Although perhaps this objection cannot be answered in a sufficiently precise sense, one plausible answer involves the claim that past tokens of a trait that currently has a function have been able to perform the activity in question over a *range* of environments. A migrating Monarch butterfly, for example, samples a range of temperatures, foliage types, and sources of nourishment. Therefore, even if the current environment of a given Monarch butterfly is, with respect to its precise specification, completely different than any that have been encountered before, nonetheless, one can say that, so long as the values of the variables that are used to define the environment fall within the range that has been established historically, then those traits that evolved in the context of migration

will be in the normal environment for their functioning. This is not to say that the answer is entirely satisfactory. Nonetheless, vagueness can only be eliminated to a certain extent. One cannot but tolerate some measure of vagueness.

Chapter 4: From Etiological Functions to Selection Processes

The previous chapter argued that etiological theories of function are uniquely capable of satisfying both adequacy conditions. However, the first section (Section 3.1) showed that there are *many* types of etiological theories, and they differ in their substantive consequences. The purpose of this chapter, then, is to evaluate the four basic etiological theories – weak persistence-based (WPE), strong persistence-based (SPE), weak reproduction-based (WRE), and strong reproduction-based (SRE), and justify the selection of the *strong persistence-based* etiological theory as particularly appropriate for the psychiatric context. According to SPE, the function of an entity consists in the activity that, in the past, contributed to the *differential persistence or reproduction* of that entity or type of entity.

The first section (Section 4.1) will clarify the logical relations between the four theories, assess some of their contrasting consequences, and argue that SPE satisfies important desiderata that a theory of function appropriate to the psychiatric context should possess. However, it will point out that other theories, particularly WRE, may be more appropriate in other scientific, and especially biological, contexts. This consequence shows that SPE is of limited value for explicating a concept of function that is applicable to the whole of biology; in other areas of biology, such as evolutionary and molecular biology and ecology, WRE (or a non-etiological theory) may be more appropriate. This conclusion reinforces the value of a pragmatic and discipline-specific approach to analyzing and refining the concept of function.

The second section (Section 4.2) will present the idea that individual learning (Section 4.2.1) and synaptic structure formation (Section 4.2.2) are mediated, to some extent, by selection processes. These selection processes are analogous to the operation of

natural selection in the evolutionary context and they ensure the differential persistence of certain cognitive or neurological structures. According to SPE, this implies that it is in some cases appropriate to assign *functions* to some learned dispositions and synaptic structures. This confirms the appropriateness and empirical relevance of SPE in the context of psychology and neuroscience and hence, by extension, to psychiatry. This section will not claim that selection processes mediate all learning or synaptic structure formation, but that the prevalence of selection processes in those realms implies that SPE is widely applicable to them.

Two conclusions will be drawn from the foregoing considerations. The first is that the empirical problem that is often raised with the etiological theory of function – that is, that there are insurmountable problems to testing function ascriptions empirically – is only a problem if "selection processes" are restricted to those that operate over an evolutionary time-scale (see Section 4.1.4 for an elaboration of this claim). The claim that selection processes mediate the formation of many learned dispositions and neural structures is empirically testable even if it is technically very difficult to test in a given case. The second conclusion is that the mere fact that a given psychological or behavioral condition is in some sense *maladaptive* in a given environment (in the sense that it reduces fitness in that environment) does not allow one to infer that that condition stems from an internal *dysfunction* (in the sense defined in the previous chapter, namely, it is unable to perform its function even in the normal environment for its functioning). This is because even learned dispositions or synaptic structures that are currently maladaptive may have been formed, in part, by selection processes, and hence they may be functioning normally relative to the environment in which they were selected, or unable to function due to an abnormal environment. This conclusion raises the level of evidence that would be required to substantiate the claim that a given psychological or behavioral

condition stems from an internal dysfunction. Chapter 5 will apply the notion of function developed here to examples from current psychiatric research on schizophrenia.

4.1 Four Types of Etiological Theory

This section will begin by providing a schematic overview of the logical relations between the four types of etiological theory (Section 4.1.1). Since all four theories that will be evaluated satisfy both of the adequacy conditions (CA_1 and CA_2*) (Section 3.3), there is no question of privileging any one of the four theories on that basis. Rather, the justification for choosing SPE will rely partly on methodological grounds, partly on grounds of intuitive plausibility, and partly on pragmatic grounds. Section 4.1.2 will illustrate the shortcoming of WPE on intuitive grounds; Section 4.1.3 will discuss the shortcoming of WRE on pragmatic grounds, and Section 4.1.4 the shortcoming of SRE on largely methodological grounds. Section 4.1.5 introduces and defends the appropriateness of SPE on all three grounds.

4.1.1 Logical Relations between the Four Theories

For the sake of convenient reference, the four theories may be schematically recapitulated as follows:

> (i) *Weak Persistence-Based Etiological Theory* (WPE): The function of an entity consists in that activity that, in the past, contributed to the persistence or reproduction of that entity or type of entity;
>
> (ii) *Weak Reproduction-Based Etiological Theory* (WRE): The function of an entity consists in that activity that, in the past, contributed to the reproduction of that entity or type of entity;

(iii) *Strong Persistence-Based Etiological Theory* (SPE): The function of an entity consists in that activity that, in the past, contributed to the differential persistence or reproduction of that entity or type of entity;

(iv) *Strong Reproduction-Based Etiological Theory* (SRE): The function of an entity consists in that activity that, in the past, contributed to the differential reproduction of that entity or type of entity.

These four theories can be arranged into a hierarchy of generality. There are two important relations to consider. First, and trivially, "reproduction-based" theories are less general than "persistence-based" theories, but this is simply an artifact of the way "persistence-based theory" is defined (see Section 3.1.2, under "Second Distinction: Reproduction-based vs. Persistence-based Theories", and especially *fn*. 74 of that chapter). It is defined disjunctively, as a theory according to which the function of an entity consists in the activity that, in the past, contributed to the (differential) *persistence or reproduction* of that entity or type of entity. Consequently, a reproduction-based theory is a type of persistence-based theory.

Second, strong etiological theories are less general than weak etiological theories. According to a strong theory, if an entity has a function then, in the past, it contributed to the *differential* persistence or reproduction of that type of entity and thereby to its own differential persistence or reproduction; according to a weak theory, if an entity has a function then it (merely) contributed to the persistence or reproduction of that type of entity. Clearly, if something contributes to the differential persistence or reproduction of

an entity it contributes to the persistence or reproduction of that entity. Consequently, a strong theory is a type of weak theory.

The following figure (Figure 4.1) illustrates the relations of generality that obtain between the four theories, where an arrow goes from one theory to another if the first is more general than the second.

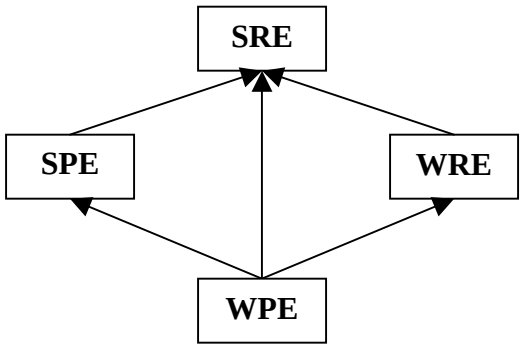

Figure 4.1: Four types of etiological theory. An arrow goes from one theory to another if the first theory is more general than the second.

How, then, can an appropriate theory of function be selected from this set of theories? As noted above, there are at least three sorts of standards that can be brought to bear on the selection of a theory of function. These consist of *methodological grounds*, *grounds of intuitive plausibility*, and *pragmatic* grounds:[108]

> (i) "methodological" grounds in this context simply refer to the very general rules or maxims that guide conceptual analysis or scientific theorizing, such as the principle that, all else being equal, simpler theories are preferable to more complicated ones, or more precise theories are preferable to less precise theories,

[108] "Pragmatic" grounds will here include the empirical adequacy and relevance of the definition of "function" for the scientific discipline it is intended to be useful for.

and so on. In the following, the methodological maxim that will find application is that, all else being equal, more general theories of function are preferable to less general theories, because the most general theory builds in the fewest potentially questionable assumptions. Hence, given that all four theories of function satisfy the conditions of adequacy that were presented in the previous two chapters, then the most general theory that satisfies those conditions should be accepted. However, the other two grounds must qualify this maxim;

(ii) "grounds of intuitive plausibility" refer to fairly coarse and very general intuitive assessments of the correctness or incorrectness of function ascriptions. For example, though WPE is the weakest theory (and hence should be privileged according to the first ground) it will be argued that WPE is *too* inclusive because it can ascribe functions to purely physical properties such as mass *as such* (e.g., in the sense of, "the function of an object's mass is to make it subject to gravity"), and it can ascribe functions to ubiquitous non-biological structures such as a pile of rocks or a fallen stick. This seems counterintuitive if the definition of function is to lend itself to a corresponding notion of "dysfunction".[109] Clearly, such intuitive judgments can be problematic and controversial, and they are subject to modification when other desiderata are brought to bear upon them. For example, philosophers used to argue on intuitive grounds that any acceptable theory of function must apply equally to artifact functions and to biological functions (e.g., Wright [1973, 143]; Boorse [1976, 77]; see Lewens [2004, 11-16] for discussion). However, today very few philosophers believe that an adequate theory of

[109] Suppose a little rock holds up a big rock in a fast-moving stream, as a consequence of which the little rock is held in place. Even if there is some concept of "function" according to which one can plausibly attribute to the little rock the "function" of holding up the big rock, it seems counterintuitive to say that the little rock becomes "dysfunctional" if the large rock falls off of it.

biological function must account for the ascription of functions to artifacts (e.g., Godfrey-Smith [1993, 347]). In other words, philosophical intuitions about the adequacy of such definitions changed over time as a consequence of the fact that other sorts of desiderata came to bear on the problem of explicating "function" – for example, that of constructing a theory of function that would be appropriate to the context of evolutionary biology.

More typically, philosophers may agree that a given sort of example is counterintuitive but disagree as to its significance. Amundson and Lauder (1994), for example, argue that despite the fact that the ascription of functions to rock formations, fallen sticks, or other ubiquitous non-biological phenomena is counterintuitive, it should not undermine the acceptance of their theory of function, which essentially amounts to an interest-based consequentialist theory that they believe to be useful in the context of evolutionary morphology (Ibid., 346-7). Since evolutionary morphologists are not interested in rock formations, then, they argue, the fact that a theory of function that morphologists find useful ascribes functions to rock formations is irrelevant. In this case, although Amundson and Lauder admit to sharing the intuitions in question, they argue that pragmatic factors (adequacy of a theory of function for a specific disciplinary context) take precedence over intuitive factors.

Nonetheless, the position that will be adopted in this dissertation is that while intuitions, taken apart from all other considerations, cannot be determinative for selecting an appropriate theory of function, they can be influential. Their clearest influence shows itself in the following consideration: of two theories of function, if one of them violates fairly common and very general intuitions about usage, and another does not, then in the absence of strong

pragmatic or methodological grounds for rejecting the latter, the latter should be accepted over the former because it more accurately reflects the usage that it is attempting to capture;

(iii) "pragmatic" grounds for accepting a theory of function are those that refer to the pragmatic goals or aims that such a theory is intended to satisfy as a basis for selecting one theory over another. The aim of the following analysis is to provide a theory of function that is appropriate to the psychiatric context, and in particular, to the cognitive, behavioral and neurobiological research that is central to that field. Clearly, disciplines such as neurobiology and cognitive science are central to contemporary psychiatry; consequently, if a theory of function permits function assignments to be made in those fields in ways that largely conform to practice, then that theory of function should be privileged over others that do not, despite the fact that the theory may be inadequate in other fields, and hence fail the methodological test of generality.

The reason for elaborating these three "grounds" for selecting a theory of function is that each has some argumentative role to play in the selection of SPE as particularly appropriate to the psychiatric context. According to the first ground, the most general theory should be selected; this methodological criterion gives preference to WPE, since, as depicted in Figure 4.1, WPE is the most general such theory. However, according to the second ground, WPE violates very general intuitions about the appropriateness of function ascriptions (it is over-inclusive) and the other three do not – or at least, not to the same extent. Therefore, on grounds of intuitive plausibility, one should reject WPE. On methodological grounds, then, one should select *either* SPE *or* WRE *over* SRE, because

SPE and WRE are the next most general theories. (This option is usually overlooked in discussions of function.) The final selection of SPE over WRE is ultimately based on pragmatic grounds.

Interestingly, the ensuing discussion will suggest a fairly intriguing division of labor: while SPE can often be found adequate for neurobiology and some aspects of cognitive psychology, it is inadequate for some aspects of evolutionary and molecular biology and ecology for which WRE is adequate, and vice versa. This is because, on the one hand, SPE can be applied to non-reproducing populations that typically undergo selection, and important aspects of synapse formation and individual learning fall under this category. On the other hand, WRE can be applied to reproducing populations that do not undergo selection, and such populations play an important role in evolutionary and molecular biology and ecology. This point will be elaborated in Section 4.1.3.

4.1.2 Inadequacy of WPE on Intuitive Grounds

As noted above (see Figure 4.1), WPE is the weakest, that is, the most general theory of function, and hence includes all of the phenomena that are captured by the other three. WPE simply entails that the function of an entity consists in that activity that, in the past, contributed to the persistence or reproduction of that entity or type of entity. This suggests that WPE should be privileged on methodological grounds because it is the most general theory that is compatible with CA_1 and CA_2*. However, it has often been argued that WPE is *too* inclusive – it awards "functions" to phenomena in ways that are intuitively implausible. The following section will first discuss the intuitive inadequacy of WPE (for reasons that have long been pointed out in the literature) and then it will examine two attempts to modify WPE in ways that exclude some of those counterintuitive consequences. The first attempt is to restrict functions to those activities that have, in the past, contributed to their own persistence *by contributing to the*

persistence of the system in which they are embedded. The second attempt is to restrict functions to those activities that have, in the past, contributed to their own persistence *in a sufficiently "complex" manner*. Both will be rejected, leaving either WRE or SPE as the next most appropriate theory of function.

Intuitive Implausibility of WPE

One way of illustrating the intuitive implausibility of WPE is to present a specific WPE view and show how it leads to the problem in question. The classic exposition of WPE is Wright (1973; 1976). To quote from this influential explication of "function":

> The function of *X* is *Z means*
> (a) *X* is there because it does *Z*,
> (b) *Z* is a consequence (or result) of *X*'s being there. (Wright [1973, 161])

In the artifact context, *X*'s form can be explained by the fact that someone recognized that form to have a certain capacity (*Z*), and produced it for that reason, thereby fulfilling the first premise. In the biological context, if *X* was selected for by virtue of one of its effects, *Z*, and this selection process partly accounts for its present existence, then it will be true to say that "*X* is there because it does *Z*", thereby also satisfying the first premise. If *X*'s being there allows it to *continue* to do *Z*, then the second will be fulfilled as well. Clearly, the purpose of Wright's fairly general analysis is to present the idea of a "cyclical"[110] causal process, one that incorporates both natural and artifact functions. His is clearly a *weak* etiological theory since there is no explicit specification that *X* must have been *selected for* doing *Z*. Moreover, it is a persistence-based theory, since his formulation is compatible with the possibility that doing *Z* merely

[110] In the sense defined in *fn*. 66.

allowed a specific token of *X* to *persist*, *or* that doing *Z* allowed one token of *X* to *reproduce*, creating a new token of *X*.

But Wright's general definition is also satisfied by processes that, intuitively, one would not want to ascribe functions to, such as the sort contrived by Boorse (1976) in his critique of Wright. Suppose, for example, that a hose in a laboratory springs a leak, and thereby emits a noxious chemical, and any scientist that attempts to seal the hose gets knocked unconscious by the chemical it emits. Thus it can be said that the leak in the hose contributes to its own persistence by knocking out anyone that comes close enough to fix it (Ibid., 72). Strictly speaking, one must say that the leak is there because it knocks out scientists, and that knocking out scientists is a consequence of its being there. But it seems counterintuitive to say that knocking out scientists is the function of the leak – again, supposing that the relevant notion of "function" can support a corresponding notion of "dysfunction". Similarly, obesity can contribute to a sedentary lifestyle, which in turn can reinforce obesity. Thus it is possible to explain a person's current obesity in terms of one of the consequences his or her obesity produced in the past that contributes to its own persistence (Ibid., 75-6). Yet, like the hose example, it is counterintuitive to suggest that the function of obesity is to contribute to a sedentary lifestyle. Bedau (1992, 786) uses an example of a stick floating down a stream that brushes against a rock and gets pinned there by the backwash it creates, and thus is responsible for perpetuating its current position, to make the same point. Clearly, trivial examples from the natural world as well as from the realm of artifacts can be multiplied indefinitely.

Boorse's counterexamples have been influential in shaping the development and refinement of etiological theories of function, since it has led many to accept that having been selected for by natural selection, rather than merely having contributed to the continuation of one's present state, is a necessary condition for having a function (see,

e.g., Neander [1983, 103]; Millikan [1993, 34-6]; Boorse himself [1976, 76] suggests this possibility but rejects it). This is tantamount to moving directly from WPE to SRE, some version of which, as noted above (see Section 3.1.2) is probably the most widely held theory of "function" amongst philosophers. Obesity, though it secures its own persistence by contributing to a sedentary lifestyle, is in no sense *selected over* some other phenotypic trait *because* it contributes to a sedentary lifestyle. Similarly, the leak in the hose is not there because *it*, rather than something else, proved to be more effective in knocking out scientists. However, the logical schema described in Figure 4.1 also shows that it is not necessary to accept SRE simply in order to avoid the problems associated with WPE. Instead, one could accept WRE or SPE instead. However, before rejecting WPE outright, one should examine the possibility that it can be modified or finessed. Below, two different attempts to modify WPE will be presented, and both of them will be shown to be inadequate to their original aim.

First Attempt: Contribution to Persistence of Containing System

Godfrey-Smith (1993, 198-199; 1994, 348-350), following Millikan (1984), restricts functions to entities that undergo some form of reproduction or replication. Consequently, he offers a reproduction-based view and not a persistence-based view. However, some of the additional restrictions he places on his theory can be applied to the persistence-based theory as well.

After restricting functions to the parts of reproducing entities, Godfrey-Smith (1994, 348) comes up with additional counterintuitive consequences for his view. Segregation-distorter (SD) genes, as noted in Section 3.1.2 (under "Two Additional Variables: System and Temporal Variables"), guarantee the overrepresentation of their chromosome in the gamete pool by "sabotaging" sperm containing the homologous chromosome. This constitutes a form of selection operating at the level of alleles.

However, Godfrey-Smith claims that it seems counterintuitive to say that disrupting meiosis should be considered the *function* of the SD genes (Ibid.). To ascribe a function to SD genes merely by virtue of the fact that they do something which contributes to their reproduction in succeeding generations would be as counterintuitive as saying that *people* have functions because they do things that contribute to their reproduction in future generations. He claims that the reason it seems counterintuitive is that the SD genes guarantee their own reproduction without contributing to the fitness of a larger system in so doing: "One way to exclude both people as bearers of functions and also exclude disruption of meiosis as a function of segregation distorters is to stipulate that (i) the functionally characterized structure must reside within a larger biologically real system, and (ii) the explanation of the selection of the functionally characterized structure must go via a positive contribution to the fitness of the larger system" (Ibid., 349). According to Godfrey-Smith, the difference between these intuitively implausible cases of function ascriptions, on the one hand, and intuitively plausible ones, on the other, is that in the latter, the functional entity contributes to the fitness of the system that contains it and thereby *indirectly* contributes to its own reproduction over time.

Even though Godfrey-Smith requires that an entity must contribute to the fitness of the larger system in order to have a function, perhaps it would suffice for saving WPE to require merely that, in order to have a function, an entity must have contributed to the *persistence* of a system of which it is a part, and *thereby* to its own persistence over time.

This relativization of function ascriptions to parts of systems would not imply that the only relevant biological "system" is the individual – i.e., the organism. Godfrey-Smith (1994, 394), like Griffiths (1993, 416) points out that selection operates on several different levels, and that functions can be assigned by virtue of the fact that an entity contributes to the reproduction of *some* containing system, even if not the organism. For

example, Godfrey-Smith (1994, 349-350) accepts that the component *parts* of SD genes can have the function of contributing to the differential reproduction of the chromosome itself.[111]

Can this additional criterion be invoked to resolve the Boorse-style counterexamples? One might be tempted to say that a stick that is pinned in place by its own backwash does not constitute a *part* of a larger system the persistence of which it contributes to, and hence it does not have a function. One might also say of obesity that, because it does not contribute to the persistence of the obese individual – in fact, to the extent that obesity is detrimental to one's health, it guarantees its own persistence *at the cost of* the persistence of the individual – then it, too, does not have a function. However, there are two problems with the use of this relativization to resolve Boorse's counterexamples. The first is that equally implausible natural examples can be derived which clearly fit the pattern of explanation that is being demanded. Godfrey-Smith (1993, 198) uses the example of a small rock that holds up a large rock in a fast-moving stream. The small rock holds up the large rock, in the absence of which the small rock would be washed away. Hence it purchases it own persistence by contributing to the persistence of a larger system of which it is a part, but, unless it functions as an artifact – i.e., someone placed it there for a reason – it seems strange to say that the little rock has "malfunctioned" if it is ultimately washed away.

More problematically, however, the part/whole distinction is always relative to a method of analysis. What one considers a "part" of a system, versus the "whole" system, depends largely upon conventional and pragmatic choices about how to analyze it. Certainly, it seems strange to say of a stick that is pinned in place by its own backwash

[111] His rationale, however, for attributing functions to the *parts* of the SD gene, and not the SD gene itself, is puzzling. SD genes ensure their overrepresentation in the gamete pool by increasing the fitness of the chromosomes that contain them. In sabotaging the homologous chromosome, then, the SD genes contribute to the fitness of a containing system, and hence should possess a function on his view.

that it is a "part" of a larger system the persistence of which it contributes to. But why should it seem strange? The structure composed of the stick, the rock that it is pinned to, and the backwash that keeps it in place, constitutes a "system" of which the stick is a "part", and the persistence of which it contributes to. By the same token, the stick itself may be seen as the "whole" system of which the parts of the stick are components, and by virtue of which each of the parts have functions. Consequently, the part/whole distinction cannot be used to draw a substantive distinction between those entities that can and those that cannot be said to possess "functions", that is, in resolving the first-level demarcation problem (see Section 3.1.3).

Second Attempt: Complex Contribution to Self-Persistence

Schlosser (1998) attempts to resolve the problem of overbreadth by restricting functions to parts of "complex self-re-producing systems" (Ibid., 305) (see Section 3.1.3, under "Goal-Contribution Theories", for a brief discussion of his view). First, his notion of "self-re-production" will be explained, and then his notion of "complexity". His attempt will then be criticized as inadequate to the task.

A system is a "self-re-producing" one if it undergoes a series of state-transitions that ensures the recurrence of certain states (Ibid., 311). For example, the earth's circling the sun, or a swinging pendulum, are fairly simply self-reproducing systems because they undergo a cyclical series of state transitions, in which a given state (e.g., position of the planet at a given time) is necessary for the next state (its position at a later time), which, in turn, is necessary for the recurrence of the original state. (Schlosser [Ibid., 305] uses the hyphenated expression "re-production" to signify the recurrence of a given state or event over time; hence, the "re-production" of a state can refer either to the intra-generational persistence of a state within an entity or the trans-generational reproduction of that state through the transmission of hereditary material from parent to offspring.

Hence, "re-production" simply means the same as what has been referred to in this dissertation as "persistence or reproduction".) He observes that traits that have "functions" are typically those that are, under some circumstances, necessary for their own re-production. For example, in certain circumstances (e.g., in an environment with predators), some feature of a trait (e.g., a pattern of wing coloration) is necessary for bringing about a certain activity (predator avoidance through camouflage). This activity, in turn, is necessary for the reproduction of that same pattern of wing-coloration in the descendents of that organism (that is, insofar as avoiding predators is necessary under some circumstances for allowing the organism to survive long enough to leave descendents that inherit that pattern of coloration). Hence, Schlosser claims that in order to have a function, a trait must be part of a "self-re-producing" system and that circumstances must exist under which it can perform an activity that is necessary for its own re-production.

Schlosser's theory of function is a *persistence-based* theory because it permits an entity to have a function merely by contributing to its own persistence. It is also a *weak* theory because it does not require that an entity must have been selected for in order to possess a function (Ibid., 323). Consequently, Schlosser endorses a weak persistence-based theory of function. Although Schlosser's theory is a consequentialist one, it could easily be transformed into an etiological theory if it were accepted that the relevant "contributions to self-re-production" are partly explanatory for why the trait in question currently exists. If one were to make this transformation then his view would constitute a version of WPE.

As a version of WPE, however, it inherits the problems that were raised above in relation to Wright's (1973) theory, namely, the Boorse-style counterexamples that trivialize the view. Schlosser recognizes, for example, that if "being necessary under

some circumstances to one's own self-re-production" were sufficient for having a function, then one would have to say that the function of the planetary orbit is to re-produce itself, or the function of the displacement of the pendulum from the vertical position is realignment with the vertical position, or the function of obesity is to contribute to a sedentary lifestyle (Schlosser [1998, 311]). Consequently, Schlosser restricts the type of systems the parts of which can have functions to *complex* self-re-producing systems. A complex self-re-producing system is one that "does not [merely] pass through simple cycles of states, but instead can re-produce a certain state via different sequences of state transitions depending on the environmental conditions" (Ibid., 312). Attributing complexity to a system, then, implies that the system must be able to exhibit some plasticity or variability in the re-production of a state. However, this complexity criterion must be analyzed more precisely in order to evaluate whether it successfully resolves the counterexamples that it is intended to resolve.

In his formal definition of "function" (Ibid., 315), complexity is defined as an attribute of the relationship between a functional trait, X, the activity, F, which constitutes its function, and the re-production of X at a later time. In other words, the system itself is not explicitly designated in the formal definition, but only the relations between the functional entity and its functional effect. According to his definition:

> F_c is a function of $X_c(t)$ iff:
> for a certain period of time $t_0 < t < t + x + y < t_0 + T$
>
> (1) $X(t)$ is directly causally necessary to establish $F(t + x)$ (under certain circumstances c_1)
> (2) $F(t + x)$ is indirectly causally necessary to establish $X(t + x + y)$ (under certain circumstances c_2)
> (3) the causal relations between $X(t)$, $F(t + x)$, $X(t + x + y)$ are complex.
> (Ibid., 315)

In the following, in order to minimize the proliferation of variables, the following convention will be utilized: X_1 refers to the presence of X at time t_1; F_2 refers to the functional activity that X_1 produces at time t_2, and X_3 refers to the recurrence (re-production) of X at time t_3, where $t_1 \leq t_2 < t_3$ and all three temporal moments fall within some interval T; c_1 refers to a circumstance under which X_1 is necessary for F_2, and c_2 refers to a circumstance under which F_2 is necessary for X_3.

"Complexity", then, describes the way in which X_1 produces F_2 or the way in which the performance of F_2 ensures the re-production of X_3. The notion of complexity will first be defined informally, then formally.

Informally, if one of these two relationships is "complex", then it must be the case that *either*: (i) under different circumstances, there are correspondingly different activities that X_1 must perform in order to produce F_2; *or* (ii) under different circumstances, there are correspondingly different consequences that F_2 must produce in order for X_3 to be re-produced. Formally, if one of these two relationships is "complex", then, it must be the case that there exists a set of different circumstances C (where $c_i \in C$, $i = [1, 2, \ldots, n]$, and $n \geq 2$), such that *either*:

(i) if $c_i \in C$ and $c_j \in C$ $(i \neq j)$ then there is an activity, F_{1i}, such that under c_i, X_1 must do F_{1i} in order to do F_2, and there is an activity F_{1j}, such that under c_j, X must do F_{1j} in order to do F_2, and F_{1j} is not the same type of activity as F_{1i}; *or*

(ii) if $c_i \in C$ and $c_j \in C$ $(i \neq j)$ then there is an activity, F_{2i}, such that under c_i, F_2 must do F_{2i} in order to do X_3, and there is an activity F_{2j}, such that under c_j, F_2 must do F_{2j} in order to do X_3, and F_{2i} is not the same type of activity as F_{2j}.

An example of a trait that exhibits complexity in performing its function comes from an examination of the mechanism of color change in cuttlefish (order Sepiida) and other Cephalopod mollusks, one of the functions of which is camouflage (crypsis) (Fogden and Fogden [1974]). Each chromatophore – pigment-filled cell – in the cuttlefish is surrounded by muscle cells, and is capable of rapid contraction and relaxation. In a relaxed state, the pigment in the cell is centrally concentrated and unobtrusive; when the muscle cells contract, the chromatophore is flattened in such a way that the pigment expands, thus permitting rapid color changes in response to environmental variation. (Since each chromatophore is separately controlled, the cuttlefish is capable of generating a vast diversity of complex cryptic patterns.) It seems clear that each chromatophore (X_1) has the function of crypsis (F_2), and since each chromatophore is, at the very least, capable of two different states – relaxation and contraction – each of which contributes to crypsis, then it can be said to perform its function in a complex manner. More precisely, there exist some circumstances, c_{1a}, under which relaxation (F_{1a}) is necessary for crypsis (F_2); there exist other circumstances, c_{1b}, under which contraction (F_{1b}) is necessary for crypsis (F_2), and there exist circumstances, c_2, under which crypsis is necessary for the reproduction of the cuttlefish and hence the reappearance of chromatophores in succeeding generations (X_3). (See Figure 4.2.)

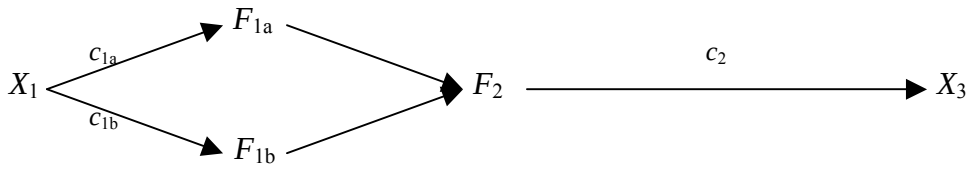

Figure 4.2: Example of complexity in the relation between X_1 and F_2. Here, there are two different circumstances, c_{1a} and c_{1b}, such that in c_{1a}, X_1 must perform F_{1a} in order to perform F_2, and in $c_{1b,}$ X_1 must perform F_{1b} in order to perform F_2. The relationship between F_2 and X_3 may itself be simple or complex; that is, there need only be a single set of circumstances c_2 under which F_2 is necessary for X_3.

However, there are other traits that plausibly have functions but that are not capable of complexly producing their functional effect. Unlike the cuttlefish, the cryptic properties bestowed by the melanic form of the peppered moth (*Biston betularia*) are produced in a fairly simple (non-complex) manner (Kettlewell [1974]); since the moth cannot change color during its lifetime, the relation between black coloration and crypsis cannot be brought about by multiple different pathways. In other words, for the moth, it is not the case that under different circumstances, there are different activities that X_1 can perform in order to produce F_2. However, even though the relation between black coloration and crypsis is not complex, the relation between crypsis (F_2) and the reproduction of black coloration in succeeding generations (X_3) may itself be complex. For example, the crypsis (F_2) afforded by black coloration enables the peppered moth to engage in one of several different activities that are necessary for the reproduction of the trait in its descendents. Under certain circumstances c_{2a} – those in which it is necessary to gather food in order to survive, for example – crypsis (F_2) allows the moth to gather food (F_{2a}) and thereby contributes the re-production of black coloration (X_3). Under other

circumstances c_{2b} – those in which it is necessary to find mates in order to reproduce, for example – crypsis (F_2) allows the moth to seek mates (F_{2b}) and thereby contributes to the re-production of black coloration (X_3). Hence the relation between the initial token X_1 of the trait, the subsequent performance of F_2, and the succeeding token X_3, is complex. (See Figure 4.3.)

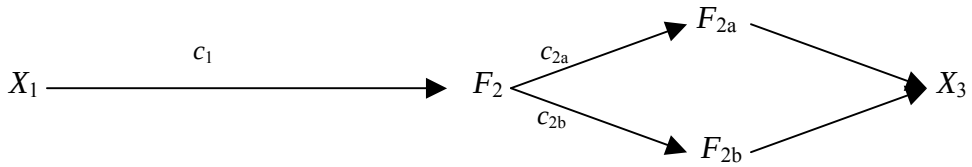

Figure 4.3: Example of complexity in the relation between F_2 and X_3. Here, there are two different circumstances, c_{2a} and c_{2b}, such that in c_{2a}, F_2 must perform F_{2a} in order for X_3 to be re-produced, and in $c_{2b,}$ F_2 must perform F_{2b} in order for X_3 to be re-produced. The relationship between X_1 and F_2 may itself be simple or complex; that is, there need only be a single set of circumstances c_1 under which X_1 is necessary for F_2.

This way of defining complexity excludes some trivial cases. For example, a non-coding segment of DNA cannot be said to have the "function" of serving as a template for its own reproduction. For although the presence of a non-coding segment of DNA at time t_1 (X_1) is necessary for it to serve as a template for its own reproduction at time t_2 (F_2), it is not the case that there exist different circumstances under which that segment must perform correspondingly different activities in order to serve as a template for its own reproduction. Moreover, even though its capacity to serve as a template for its own reproduction at t_2 (F_2) is necessary for its reappearance in descendent organisms at time t_3 (X_3), it is not the case that there exist different circumstances under which F_2 must produce correspondingly different consequences in order to ensure its subsequent

appearance in future generations (X_3). The only activity that F_2 allows it to perform is entering into a gamete that will ensure its intergenerational transmission. There is not an additional set of circumstances under which, for example, it specifies a protein that performs a significant role in the development of the organism and thereby contributes, indirectly, to its intergenerational reproduction.[112]

On similar grounds, one could not say of a stick that is pinned to a rock because of the backwash it creates that it exhibits "complexity" in creating that backwash and hence perpetuating its own position. In order to claim that the function of the stick's position is to create a backwash, one would have to say either that there are different circumstances under which the stick must perform correspondingly different activities for creating a backwash, or that there are different circumstances under which the presence of backwash will produce correspondingly different consequences that ensure the persistence of the stick. But one can say neither of those things. The stick "re-produces" its position in a simple and non-complex way.[113]

However, the restriction of functions to complex self-re-producing systems is not sufficient for guaranteeing the appropriateness of function ascriptions, since there are two types of counterexamples to his view. The first are counterexamples that satisfy the scheme described in Figure 4.2. Although Schlosser states that his definition of complexity excludes obesity from having the function of producing a sedentary lifestyle (Ibid., 319), it is unclear why one would consider the cyclical relationship between obesity and a sedentary lifestyle to be "non-complex" in his sense. On the one hand, obesity contributes to fatigue, which contributes to a sedentary condition. On the other

[112] What is referred to as "junk DNA" – the vast stretches of non-protein-coding DNA segments once thought to be without function – is now suspected to code for RNA segments with crucial regulatory functions (e.g., Mattick [2004]).

[113] This claim depends, of course, on how finely the situation is analyzed. To this extent, the relation between complexity and simplicity is relative to a method of analysis, like the relation between parts and wholes. This point will be raised at the end of the section.

hand, obese people often feel anxious about the prospect of strenuous activity and hence devise various means to avoid engaging in it. Thus, the relation between obesity and a sedentary lifestyle can be wrought by anxiety in addition to fatigue. Hence there are different circumstances under which obesity (X_1) produces a sedentary lifestyle (F_2): under some circumstances, c_{1a}, obesity is necessary for the production of fatigue (F_{1a}), under other circumstances, c_{1b}, obesity is necessary for the production of anxiety (F_{1b}), and both fatigue and anxiety contribute to a sedentary lifestyle. In turn, a sedentary lifestyle reinforces obesity. According to Schlosser's view, then, obesity has the function of producing a sedentary lifestyle.

A well-documented psychiatric example of such a "complex self-reproducing system" involves the complex cyclical relation between panic, on the one hand, and mistaken beliefs about bodily sensations, on the other, although it would be both counterintuitive, as well as contrary to psychiatric usage, to suggest that panic has the function of producing mistaken beliefs about bodily sensations. One theory of panic stems from the realm of cognitive behavior therapy. According to this theory (Clark [1986; 1997]) a panic attack is typically initiated by a stimulus that is perceived as threatening. This gives rise to a state of apprehension, which, in turn, gives rise to bodily sensations (such as increased heart rate or dizziness). These bodily sensations are then *misinterpreted* in a "catastrophic" fashion, that is, as a signal of immanent bodily danger (heart attack, death, etc.), which incites a full-blown panic response – a sudden onset of intense fear and discomfort. Although panic attacks are not uncommon – according to community surveys, between 7 and 28 percent of the general population will experience at least one panic attack (Clark [1997, 126]) – it is much less common for a person to experience recurrent panic attacks. In the latter case, the condition constitutes "panic

disorder" (APA [2000, 433-441]) which affects about 3-5 percent of the general population (Clark [1997, 126]; see Wittchen and Essau [1991]).

There are at least two cognitive mechanisms that appear to contribute to the recurrence of panic attacks (Clark [1997, 125]). First, people who have experienced a panic attack may become frightened of re-experiencing the bodily sensations which have been misinterpreted, and as a consequence become more vigilant in monitoring bodily sensations. This has the consequence of allowing them to notice sensations of which they were previously unaware, and hence can magnify the potential for a renewed attack. Second, people with panic disorder may systematically avoid situations that induce the sensations believed to be linked with the (imagined) bodily danger. For example, they may avoid strenuous activity out of the fear of re-experiencing sensations believed to be associated with imminent danger. One consequence of this, however, is that the person avoids precisely those situations that would have the effect of *disconfirming* the false beliefs about his or her condition. For example, if a person avoids jogging because he or she comes to believe that jogging may bring about a heart attack, then he deprives himself of many occasions to realize that it will not do so (Ibid.; also see Salkovskis [1991]).

The recurrence of panic attacks, then, illustrates succinctly Schlosser's (1998) concept of "complex self-re-production"; in fact, it fits the model outlined in Figure 4.2. Under certain circumstances, a panic attack can bring about hypervigilance to one's own bodily sensations, which, in turn, reinforces one's mistaken beliefs and hence increases the probability of a renewed panic attack. In other circumstances, a panic attack can bring about avoidant behavior, which has the effect of preventing the disconfirmation of the mistaken beliefs, thereby reinforcing them and preparing the person for future attacks. Hence, according to Schlosser's view, panic attacks have the function of reinforcing

mistaken beliefs that are necessary for its re-production. This is counterintuitive, as well as contrary to psychiatric usage.[114]

The second type of counterexample involves systems that satisfy the scheme for complexity described in Figure 4.3 but that do not have functions. It was noted above that by providing crypsis, black coloration enables the peppered moth to engage in any one of a large number of different activities that are necessary for its survival and reproduction. But other properties – such as the property of having *mass*, or *solidity*, or *coloration*, also fit this scheme. Clearly, the possession of mass subjects the organism to gravity, which is necessary, under different circumstances, for the organism to be able to engage in any one of the diverse activities that lead to its survival or reproduction. But it seems counterintuitive to say that these purely physical properties have the function of doing so. As Williams (1966) points out, the ability of a flying fish to return to water is necessary

[114] Schlosser (pers. comm.) denies that the example of obesity is a serious problem for his view. In his view, the appropriateness of function ascriptions is a matter of degree, and is proportional to the number of different pathways by which a trait can contribute to its own re-production. Hence, the fact that the relation between obesity and a sedentary lifestyle can be mediated by *two* series of state transitions rather than merely *one* makes it only slightly less trivial to ascribe a function to obesity than it would be to ascribe a function to junk DNA, which can only reproduce itself by a single pathway: "Trivial cycles only become slightly less trivial if they are supported by two alternative trajectories than merely one [since] they are still a far cry from a living system with its exuberant complexity" (Ibid.). Moreover, he adds that since living systems, with their exuberant complexity, have evolved from simpler, non-living systems, one should not expect an unambiguous dividing line to separate appropriate from inappropriate function ascriptions (Ibid.) Clearly, the adequacy of Schlosser's response hinges on the assumption that there are no more than two, or at least a very few, circumstances, in which, e.g., obesity leads to a sedentary lifestyle or panic to mistaken beliefs. However, the relation between obesity and a sedentary lifestyle can be as complex as the techniques of avoidance that the human mind is capable of contriving. Under one circumstance (e.g., embarrassment about physical appearance), obesity is necessary for producing aversion to, say, purchasing a membership at a gym; under another (e.g., fear of excessive perspiration), obesity is necessary for producing aversion to, e.g., staying outdoors for long periods of time; under a third (e.g., production of fatigue), obesity is necessary to bringing about the early cessation of strenuous physical activity, and so on. Clearly, one could continue to generate such scenarios, and hence an arbitrarily large number of independent trajectories can mediate the relation between the two, making the relation between obesity and a sedentary lifestyle complex to an arbitrarily large degree. The same can be said of the relation between panic and mistaken beliefs. Moreover, if, according to Schlosser, one's warrant for asserting a function statement is proportional to the complexity of the function relationship, then one would have greater warrant for saying that the function of obesity is to contribute to a sedentary lifestyle than that the function of black coloration on *Biston beularia* has the function of predator avoidance, since, as shown above, the relationship between the latter two events can be viewed as a relatively simple one.

for its survival, but functions are not typically assigned to properties that "achieve the mechanically inevitable" (Ibid., 11-12). Consequently, imposing the "complexity" criterion alone does not appear to be sufficient for making the substantive distinction between the sorts of entities that can have functions and those that cannot.[115]

Another way of formulating this criticism of Schlosser's theory is to point out that the distinction between complex and non-complex systems – like that between parts and wholes – is always relative to a method of analysis. Consequently, whether or not a given trait possesses a "function" comes to depend, on his view, upon the level of detail with which one describes the phenomenon in question. Hence, like Godfrey-Smith's (1993; 1994) attempt, it cannot be used to draw a substantive distinction between those entities that can and those that cannot possess functions.

Perhaps other modifications could be imposed in order to produce an adequate version of WPE; however, at this point it may be useful to explore the other three etiological theories. Although, as noted above, the failure of WPE led many philosophers to accept SRE, there are two other, commonly-neglected, options available, WRE and SPE. In the following three subsections, the rationales that motivate WRE, SPE, and SRE, as well as the strengths and weaknesses of each view, will be assessed.

4.1.3 Inadequacy of WRE on Pragmatic Grounds

One solution that has been proposed to the problem of triviality is that in order for an entity to have a function, it is not sufficient that it contributes to its own persistence, but that it contributes to its own *reproduction* by contributing to the persistence or reproduction of the system that contains it (Millikan [1993, 32-35]; Godfrey Smith [1993,

[115] Schlosser (pers. comm.) points out that, e.g., the property of having a *particular* mass can have a function, but that is not the point at issue. Rather, the problem is that the rationale that leads to ascribing a function to a *particular* mass can also be used to ascribe a function to mass *as such*, which is counterintuitive.

198-199]; [1994, 348-350]). This move is tantamount to making the transition from WPE theory to WRE (See Figure 4.4.).

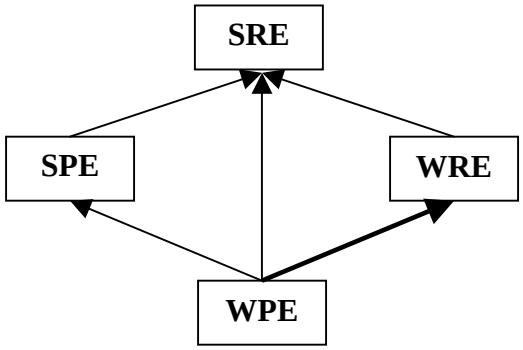

Figure 4.4: Four types of etiological theory. The thick arrow depicts the transition from WPE to WRE.

Godfrey-Smith (1993, 198-199), as noted above, in a discussion of the Boorse-style counterexamples, argues that the reason that the structures in question – such as the little rock that holds up the big rock – do not have functions is because they do not contribute to the *reproduction* of the containing system. By holding up the large rock, the small rock is itself held in place; but this merely contributes to the indefinite persistence of that structure itself, rather than the multiplication of like structures. Hence, Godfrey-Smith accepts Millikan's (1984) view that functions can only be defined for entities that exist within "reproductively established families" (Ibid., 18) – roughly, for entities that can be called, in a very broad sense of the term, "copies" of one another, or that are produced from such copies.[116] For Millikan, to say that an individual, *Y*, is a reproduction or "copy" of an individual, *X*, is to say that, (i) *X* and *Y* share some property, *P*; (ii) there

[116] Millikan (1984) clearly intends this notion of a "reproductively established family" (Ibid., 18) to be understood broadly to apply to cultural as well as biological forms of transmission: "Tokens of a specific gene, handshakes occurring in a culture (i.e., not independently derived), household screwdrivers (the same design has been copied over and over) and various tokens of the same word are members of first-order reproductively established families", as are sentences that exemplify "the same syntactic form" (Ibid., 23).

is a set of natural laws, or laws derivable from those natural laws, that entail that if *X* had differed in certain ways with respect to *P*, *Y* would have differed with respect to *P*, and (iii) *X*'s having *P causes Y* to have *P*.[117]

Although, in the initial presentation of her view, Millikan (1984) does not discuss Boorse's counterexamples, and hence does not motivate her theory of function on that basis, in a later presentation she exploits Boorse's counterexamples in order to motivate SRE. Millikan (1993) begins her discussion of function by tentatively accepting Wright's WPE view, and observing that "anything that cycles will fit [Wright's view]" (Ibid., 33). For example, "stages of an idling motor; the positions, vector velocities, and accelerations of the planets; the various stages of the earth's water cycle; and so forth will have functions by this definition" (Ibid., 34). To resolve these problems, she introduces the restriction that the functional entity must be part of a reproductively established family; that is, the production of the entity must have involved "reproduction or copying of its functional features" (Millikan [1993, 34]). By restricting function ascriptions to entities that are capable of reproduction, Millikan implicitly (and, as will be seen, only tentatively) accepts a version of WRE.

Buller (1998; 2002) is also led to endorse a version of WRE; however, rather than reaching it by strengthening WPE to exclude counterexamples, he weakens SRE in a way that he thinks preserves the core insight of the etiological view but that does away with some unnecessary restrictions that are associated with SRE (see discussion of Buller's [1998] view in Section 3.1.2, under "First Distinction: Weak vs. Strong Etiological Theories"). He correctly points out that the rationale for SRE is that it supports the intuition that to attribute a function, *F*, to a trait, *T*, of an organism, *O*, is, in part, to

[117] Presumably, clause (iii) of the definition is intended to exclude *Y* from being a "copy" of *X* if *X* is like *Y* because they were both molded from a common template, or if *Y* had *P* before *X* had *P*. For example, two items from the same assembly line are not "copies" of one another but of the same template (Ibid., 21), even though the two items would satisfy clauses (i) and (ii) of the definition.

explain why *T* currently exists. More precisely, it is to say that *T*'s doing *F* in organisms like *O* figures into a complete explanation for the current presence or distribution of *T*s in the population of organisms. If *T* has been selected for because of its capacity to do *F*, then clearly, *T*'s doing *F* will figure into a complete explanation for why *T* is there. However, Buller goes on to argue that as long as ancestral tokens of *T* did *F*, and *T*'s doing *F* contributed to the survival and reproductive capacity of *O*'s ancestors, and *T* is heritable, then the fact that ancestral tokens of *T* did *F* must figure into a complete explanation for why *O* has *T*, even if *T* was never selected for. It does not appear that reference to selection, that is, the differential fitness bestowed upon *O*'s ancestors by *T*, is required of a theory of function in order for it to possess the same type of explanatory power possessed by SRE.

There are at least three advantages to accepting WRE. First, in addition to preserving the explanatory power of SRE, WRE avoids many possible counterexamples, in addition to those of Boorse. For example, as noted in the previous section (Section 4.1.2), the property of possessing mass *as such* is not the sort of property that one attributes a function to. Nonetheless, an organism's possessing mass is required in order for it to perform any of the activities required for survival and reproduction. Consequently, if all that is required for a property to have a function is that that property contributed in some manner to an organism's reproductive capacity, and thus ensured its own reproduction in future generations, then it seems that one would have to attribute a function to mass. However, Buller (1998) points out that in order for a trait to be hereditary, it must be capable of variation within a population. Since organisms do not vary with respect to whether or not they possess mass, then mass is not hereditary and hence does not have a function (Ibid., 517-518) – although, as noted in *fn*. 115, one might attribute a function to having a *certain* mass rather than another.

Second, as noted earlier, WRE allows traits that were not selected for to have functions, and this is clearly more consistent with biological usage than the view that the function of an entity consists in a selected effect. Since an entity need not have undergone selection in order to have a function, then functions can be awarded to traits that, for example, have gone to fixation by virtue of genetic drift or that are the inevitable consequences of a developmental constraint, but that nonetheless contribute to their own reproduction.

Third – though this point is actually a generalization of the second point – WRE permits the assignment of functions in biological contexts where selection is not operative at all. For example, the "functional approach" to ecology examines the ecosystem roles that are performed by groups of organisms (Cummins [1988, 254]). These "functional groups" are usually composed of diverse biological taxa that have shared characteristics, such as morphological and behavioral mechanisms of food acquisition (e.g., leaf shredding, particle filtration, etc.), or a shared position within a trophic structure (e.g., autotrophs, decomposers, and consumers) (Naeem and Li [1997, 508]). The shared characteristics that define a functional group are identified by virtue of their role in maintaining some complex ecosystem property, such as ecosystem stability or resilience. Such groups are said to have a "function", even though there is no clear sense in which one functional group of organisms has been "selected for" this activity over another group. Pâslaru (2005) uses this point to argue against the applicability of the etiological conception of function in the ecological context and to endorse the appropriateness of Cummins' contribution-based account instead.

However, the contribution of a functional group to an ecosystem property is often only the first stage in a cyclical process that concludes by supplying the functional group itself with the material it requires for its own reproduction. For example, the carbon cycle

involves the decomposition of cellulose by microorganisms such as the bacterium genus *Streptomyces*, which produces CO_2; once produced, CO_2 is absorbed by primary producers, which are consumed and which, in turn, produce cellulose for *Streptomyces* (Meyer [1993]). Hence not only does *Streptomyces* contribute to a complex capacity of an ecosystem, but in doing so it contributes to its own reproduction. This constitutes a "consequence-etiology" in Wright's (1976, 116) sense, namely, in that an activity of *Streptomyces* figures into a complete account of its own continued presence in an ecosystem. Furthermore, since the production of CO_2 contributes to the *reproduction* of the *Streptomyces* in addition to its persistence, then WRE allows the assignment of a function to an individual *Streptomyces* bacterium itself.[118]

Moreover, as an etiological theory, WRE easily allows function ascriptions in ecology to possess normative content, in that it allows parts of ecosystems to malfunction or fail to function. That function statements in ecology may lend themselves to normative usage is apparent from Meyer's remark that microorganisms can *fail* to function: "Human, animal, and plant life on Earth would soon come to an end if the physiological groupings of microorganisms…did not function properly or became extinct" (Ibid., 68). According to WRE, since producing CO_2 is what *Streptomyces* did in the past that accounts for its own continued existence, it has this as a function even if it is not currently capable of performing that function, and hence it is capable of malfunctioning.

---

[118] This function ascription, however, may raise a potential problem of ambiguity if one is not careful to specify the *system-level* in relation to which the function is being identified. For, on the one hand, one can say that the function of *Streptomyces*' cellulose-digesting exoenzyme is the breakdown of polymers into smaller units capable of digestion by the microorganism. This is the function of the exoenzyme relative to the individual bacterium. On the other hand, one can impute the function of cellulose breakdown to the bacterium *itself*, relative to the ecosystem as a whole. As Meyer (1993) states, "The function of microorganisms…is to break down the bodies of plants and animals into simpler substances" (Ibid., 90); he adds that this contributes to the functioning of the ecosystem, thus clearly relativizing the function to the ecosystem: "A main function of microbial activity thus is to complete the cycle of chemicals through individual ecosystems and the ecosphere as a whole" (Ibid., 90-91). Consequently, the same activity – breakdown of polymers into monomers – can have a function relative to the organism, the ecosystem, or both.

By contrast, Pâslaru's (2005) purely consequentialist attempt to account for the normative role of function statements within ecology is relatively strained; he defines "malfunction" in terms of a similarity relation between a malfunctioning and functioning item.[119] Consequently, Pâslaru's concept of "malfunction" is an externalist one, because whether an entity is malfunctioning or not depends upon the availability or presence of functional tokens that are sufficiently similar to it.

WRE, however, has an important disadvantage which entails that it must be rejected in this context. WRE does not assign functions to entities that are not capable of reproduction. As noted in Section 3.1.2 (under "Second Distinction: Reproduction-based vs. Persistence-based Theories"), one of the main advantages of SPE is that it applies to complex adapted systems that are neither heritable nor capable of reproduction, such as the unique synaptic structure of the brain of a mature individual. WRE also does not lend itself easily to the ascription of functions to learned dispositions, while SPE does. One, albeit limited, behaviorist model of learning is that learning consists in the *differential reinforcement* of behavioral dispositions owing to the different consequences that they produce. Consequently, SPE can assign functions to these sorts of learned dispositions and WRE cannot, since dispositions do not "reproduce" within an individual. These points will be raised below (Section 4.1.5) to argue for the appropriateness of SPE in the psychiatric context. Consequently, to accept WRE as an adequate normative theory of function would not permit functions to be assigned to any but reproducible traits, and this would exclude a large range of possible cases in the neuroscientific and psychological contexts.

---

[119] Roughly, according to Pâslaru's view, an entity, *x*, within an ecosystem is *malfunctioning* when it is structurally incapable of performing some activity, *A*, which some other entity, *y*, is able to perform; where *y* has a "pattern of previous performances" of *A*; and where *x* is similar to *y* but for this structural divergence that accounts for its inability.

On these grounds, there appears to be an impasse, assuming that one is committed to an etiological theory of function. On the one hand, one can accept WRE on the grounds of its appropriateness to certain biological contexts such as evolutionary biology and ecology, and reject SPE. In this case, as pointed out by Godfrey-Smith (1993, 199), one excludes many functions from the neurobiological and psychological realms. On the other hand, one can accept SPE on the grounds of its appropriateness for certain neurobiological and psychological contexts, and reject WRE. But in the latter case, the function of an item consists in a *selected effect*, and consequently, if a trait has not undergone selection of *any* sort than it cannot have a function. But that is clearly inadequate to the context of evolutionary and molecular biology as well as ecology, as noted above. This suggests the possibility that a unified *etiological* theory of function might not be possible or useful even in the field of biology, and the final choice between SPE and WRE will have to be made on pragmatic and discipline-specific grounds.[120]

4.1.4 Inadequacy of SRE on Methodological Grounds

According to SRE, in order for an entity to have a function, it must have contributed to the differential persistence or reproduction of that entity or type of entity. This, as noted above (see Section 3.1.2, under "First Distinction: Weak vs. Strong Etiological Theories") is probably the most widely-held theory of function amongst

[120] One could argue that this disunity constitutes a good reason for rejecting the etiological theory altogether. However, two points can be raised against this rejection. The first is that the same plurality affects consequentialist theories as well. For example, if one accepts a consequentialist theory that identifies the function of a system part with the activity that contributes to the reproduction of the whole system, then it is not clear that one can assign functions to ecosystems or impute functions to entities that do not undergo reproduction. On the other hand, if one merely imputes functions to anything that contributes to the persistence of a system, then one cannot resolve the Boorse-style counterexamples raised earlier. These problems are by no means specific to the etiological view. The second point is that, as noted in Sections 3.2 and 3.3, etiological theories are uniquely capable of satisfying both adequacy conditions for a theory of function that permits the construction of a non-externalist definition of "dysfunction", and non-etiological theories are not. Thus, even if the theory of function that is eventually accepted as appropriate for the psychiatric context is *not* consistent with the majority of biological usage, then this fact would only go to show that the explanatory and inferential context of psychiatry is sufficiently distinct that a theory of function suitable to this context will not be suitable to all biological contexts.

philosophers. Although the paradigm case of a process that satisfies SRE is natural selection operating over a population of reproducing *organisms*, in principle, the theory can allow any selection process that operates over a population of reproducing entities to bestow functions onto those entities. The purpose of this section is to show that SRE is unnecessarily restrictive, and that the arguments that have been adduced in its favor are unconvincing. Moreover, in the absence of any convincing arguments for accepting such a restrictive theory of function, it must be rejected on the methodological ground that the most general etiological theory that satisfies the adequacy conditions CA_1 and CA_2*, and that is consistent with the other two grounds, should be accepted.

The fact that one can accept an etiological theory of function while rejecting SRE undermines one criticism that has been made against etiological theories of function in general, which stems from the observation that claims about the selective history of a given trait, and more generally, about the evolutionary history of that trait, are often very difficult to establish empirically (Gould and Lewontin [1979]).[121] This is particularly true when the claim in question concerns the evolutionary history of psychological traits, or even of their neurobiological correlates (Lewontin [1998]). Consequently, if, in order to assign a "function" to a given trait, one must establish that the trait has been selected for because of one of the activities that it produced in the evolutionary past, then many claims about the functions of traits would be almost impossible to establish empirically, and hence they would be irrelevant to much of actual scientific practice (Amundson and Lauder [1994, 356-61]; Schlosser [1998, 323-4]; Wouters [2005b, 144]). This epistemological problem has often been raised, in particular, against Wakefield's (1992;

[121] When selection acts on sub-organismic levels, for example, in immunological selection, synaptic selection, or some forms of learning that can be modeled in terms of selection, claims about the selective history of a trait lend themselves much more easily to empirical testing. Hence this criticism cannot be applied to all types of selection processes, but only to those that appeal to the operation of selection over an evolutionary time-frame.

1993; 1999a) attempt to define a notion of "mental disorder" on the basis of an etiological concept of "dysfunction".[122] However, since having been selected for by natural selection over an evolutionary time-frame is not a necessary condition on etiological function ascriptions, this criticism itself does not hold.[123]

As noted above, Millikan (1993) invokes Boorse's counterexamples for the purpose of rejecting Wright's (1973) WPE view, and she shows how restricting function ascriptions to reproducing entities excludes many of those systems from having "functions". Thus, as pointed out above, she formally moves from a WPE view to a WRE view (see Figure 4.4). However, she does not offer an unqualified endorsement of WRE; in fact, she rejects it in favor of SRE. This is tantamount to making two transitions: one from WPE to WRE and the next from WRE to SRE (see Figure 4.5).

---

[122] See McNally (1994, 205); Lilienfeld and Marino (1995, 413); Sadler and Agich (1995, 226-7); Sadler (1999, 435); Kirmayer and Young (1999, 449); Woolfolk (1999, 660); Bolton (2001, 198-9); Murphy and Woolfolk (2001, 245).

[123] Interestingly, proponents of "evolutionary psychiatry" – a branch of evolutionary psychology that deals specifically with the evolutionary background of mental disorders – often claim that mental disorders represent the *"normal" functioning of evolved mechanisms that have been placed in abnormal environments*, rather than the *failure of evolved mechanisms to perform their selected functions* (e.g., Nesse and Williams [1994, 209]; Stevens and Price [1996, 35-38]; Nesse and Williams [1997, 2-3]; Cosmides and Tooby [1999, 453-454]). In other words, there exists a "mismatch" between the environment in which human mental capacities evolved and the dictates of modern society, rather than a biological dysfunction or purely internal breakdown of evolved mental faculties. Depression, for example, is seen by evolutionary psychiatrists as an adapted response to being the "loser" of within-group bids for power, but one that is no longer adaptive in today's competitive workplace (Price *et al*. [1994]). Anxiety is seen, to a large extent, as an adapted response to a hostile and predatory environment that no longer exists (Marks and Nesse [1994]). Mental disorder, so the idea goes, is an inevitable byproduct of the too-rapid emergence of modern society with its novel selection pressures. Like the epistemological argument, this "mismatch" argument has often been brought to bear against the attempt to define mental disorder in evolutionary terms (see Lilienfeld and Marino [1995, 416]; Lilienfeld and Marino [1999, 408-409]; Richters and Hinshaw [1999, 442]; Woolfolk [1999, 662]; Bolton [2000, 148]; Bolton [2001, 194]; Murphy and Stich [2000, 81-84]; Murphy and Woolfolk [2001, 244]). However, this dissertation takes no position on the evolutionary context of mental disorders, and remains agnostic about the evolutionary history of human mental faculties. However, one important consequence of the "mismatch" argument, which will be elaborated in the conclusion of this chapter, is that one cannot reliably infer from the fact that a psychological or behavioral condition is currently *maladaptive*, that it stems from an inner mechanism that is *dysfunctional* or *malfunctioning*.

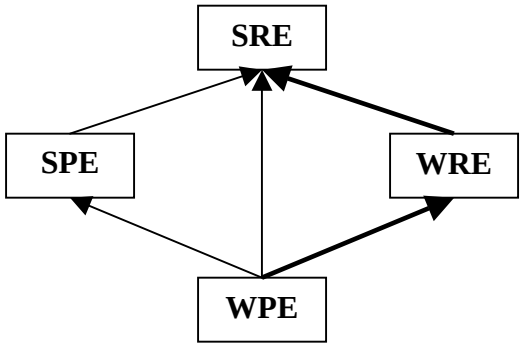

Figure 4.5: Four types of etiological theory. The thick arrows depict the transition from WPE to WRE, and then from WRE to SRE. This ignores the logical possibility of SPE.

Millikan (1993) adduces two arguments for the superiority of SRE over WRE. The first argument is that, unless some reference is made to the selective history of the functional entity itself – that is, unless the entity in question contributed not only to its own reproduction, but to its *differential* reproduction – then some of the Boorse-style counterexamples cannot be excluded. She claims that the positions, vector velocities, and accelerations of the planets *are* copied from one another and hence would satisfy WRE:

> Within quite a large latitude, had the earth been in a different place going at a different speed in a different direction last year, this placement and vector velocity would have reproduced itself this year…Simply cycling through reproductions, no matter how complex the cycle, doesn't intuitively seem to be having a function in any reasonable sense. (Ibid., 35)

In short, her argument against WRE requires that, e.g., the position of the earth at one point in its revolution qualifies as a property that is *copied* or *reproduced* from its position exactly one year earlier, and hence that WRE must assign a function to it. But

this conclusion is not required by her definition of a "copy", and in fact, it only emerges from a rather strained and artificial interpretation of that definition. According to that definition (see Section 4.1.3), the position of the earth would presumably be the property, *P*, that is reproduced from year to year; hence, the earth itself at a given moment would be the individual, *X*, which is a copy or reproduction of the earth at an earlier moment, *Y*. In other words, in order for her counterexample to work, the *earth* would have to be a "copy" of itself, and not the position of the earth from one year to the next. This is an unacceptable consequence for a definition of "copy". One fairly obvious way of avoiding this problem would be to specify explicitly that in order for an entity, *X*, to be a "copy" of an entity, *Y*, *X* and *Y* must be different entities. On this reading, the position of the earth is not a property that can have a function because it is not a property of something that undergoes reproduction.

Her second argument against WRE involves an appeal to the vagueness that she believes afflicts WRE and not SRE. Suppose one accepts WRE. Then, clearly, in order for, e.g., a trait, *T*, of an organism, *O*, to have a function, *F*, it is not necessary that in *every* single one of *O*'s ancestors that had *T*, *T* must have performed *F*. This would exclude virtually every trait from having a function unless that trait has been, historically, absolutely required for the reproduction of each member of that lineage. Yet for *T* in *O* to have a function, it is also clearly not sufficient that in only *one* of *O*'s ancestors that had *T*, *T* performed *F*. This would bestow functions upon every heritable trait that has, at some point or another, performed a useful effect – for example, one would have to assign the function of holding up glasses to the human nose.[124] Clearly, one must implicitly draw a dividing line between all and one. But how can it be non-arbitrarily drawn?

[124] Note that while it may seem appropriate to say that the nose *functions as* an eyeglasses-holder, it does not seem appropriate to say that a *function* of the nose is to hold up glasses. As Wright (1973, 147) points out, there appears to be a difference between attributing a *function* to an entity and claiming that an entity *functions as* something: the latter admits of much broader usage.

Millikan argues that natural selection can provide a non-arbitrary discrimination: "The place the line is naturally drawn…is where natural selection draws it. Only if an item or trait has been *selected* for reproduction, *as over against other traits*, *because* it sometimes has a certain effect does that effect have a function" (Ibid., 36). But this conclusion is simply mistaken. As noted in Section 3.2.1 (under "Statistical Account of Function", in particular, *fn*. 102 of that chapter), a similar vagueness afflicts the concept of having been *selected for* itself. If, in a given generation, *n*, there is at least one organism, *O*, having a heritable trait *T*, and at least one organism, *O**, that does not have *T*, and having *T* allowed *O* to leave more offspring than *O** in that generation, should *T* count as having "undergone selection" for doing *F*? Invoking *differential* reproduction does not provide the non-arbitrary dividing line that Millikan expects it to. Consequently, to the extent that the elimination of vagueness is an important *desideratum* for a theory of function to possess, then all etiological theories of function that have been examined fall short of this *desideratum*. It may be more reasonable, then, simply to reject the view that a theory of function must eliminate all sources of vagueness.[125]

---

[125] Godfrey-Smith (1993; 1994) also embraces a version of SRE, but for reasons that are less clear, and which, in fact, run counter to his express intent. As noted above (Section 4.1.3), he notes that Boorse's counterexamples afflict WPE, and he resolves them by endorsing WRE. He then simply endorses SRE without any further argumentation. For example, after endorsing Millikan's (1984) restriction of functions to entities that are part of "reproductively established families", he points out that Millikan herself requires not only that the entity is part of a reproductively established family, but that the entity must have been selected for (1993, 199). He does not criticize this view nor provide any motivation for it; shortly thereafter, he endorses the view, at least for the purpose of discussion: "In this article I will assume that an explicit appeal to selection processes and reproductively established lineages is appropriate [for assigning functions to entities]" (Ibid.). In an article written in the following year, Godfrey-Smith (1994) draws attention to Boorse's counterexamples, tentatively accepts WRE, and then points out that intuitively, one would not want to ascribe functions to people, to the chromosomes that carry SD genes, or to other entities that did not, in the past, contribute to the *fitness* of a larger system (1994, 349; also see Section 3.1.2, under "Two Additional Variables: System and Temporal Variables" for a discussion of this point). He goes on to equate fitness not merely with the reproductive propensity of that larger system (e.g., expected number of offspring) but the *relative* reproductive propensity of that system. This implies that if a trait possesses a function then it was selected for (Ibid., 350). But this view is unnecessarily strong. As noted above (Section 4.1.2), one could prohibit the ascription of functions to, e.g., the whole organism (considered as an autonomous unit) by restricting function ascriptions to parts of systems that merely contributed to the *reproduction* of some containing system and not to the *differential reproduction* of that system. Like Millikan's theory, this move altogether ignores the logical possibility of SPE or WRE.

As an etiological theory, then, SRE should be rejected on methodological grounds. Restriction of functions to the outcome of selection processes operating over reproducing entities is not necessary for preserving the normative and explanatory content of function ascriptions.

Before moving on to discuss SPE, one final argument against SRE should be noted. So far, the argument in this section has been that selection operating over a population of reproducing entities is not *necessary* for the ascription of functions to entities, though it has been assumed to be *sufficient*. Bedau, however (1991; 1993) has argued that selection is *not* sufficient for assigning functions to entities because it faces an insurmountable problem of overbreadth. But clearly, if SRE suffers overbreadth, then all of the other etiological theories must suffer overbreadth, since, as the least general theory of function (see Figure 4.1), all functions that SRE assigns are also assigned by the other three theories. Hence, if one accepts this criticism then one must reject all four etiological theories as insufficient for defining "function".

Bedau (1991; 1993) uses the example of crystal growth and reproduction to provide what he takes to be a counterexample to SRE. He argues that clay crystals undergo a form of natural selection – characterized by heritable variation associated with differential fitness – but that it seems counterintuitive to assign functions to the parts of crystals. Clay crystals, he points out, are composed of ordered layers of molecules, and crystal growth consists in the addition of new layers. When they reach a certain size, crystals cleave and break into smaller pieces that, in turn, continue to grow through the addition of new layers of molecules. Hence, crystals can be said to undergo a type of "reproduction". Environmental conditions can affect the structure of a crystal's molecular lattice, creating new variations on this structure. Since new layers of crystals tend to reproduce the geometrical arrangement of earlier layers, these variations tend to be

propagated throughout the growth process. When the crystal cleaves into smaller pieces, these smaller pieces may retain the variant structures, which therefore continue to propagate. Hence, crystal reproduction also exhibits a type of "inheritance". Most importantly, variation in the molecular lattice structure can affect many other physical characteristics of the crystal, such as its shape, growth rate, density, and cleavage conditions. These physical characteristics affect the rate at which crystals possessing a variant structure proliferate and disperse in a population; hence, one can assign a "fitness" value to different such structures. Populations of crystals, then, undergo natural selection, since such populations possess the characteristic of hereditary variation that is correlated with differential fitness (Lewontin [1970, 1]). However, Bedau argues that it is counterintuitive to ascribe *functions* to the parts of crystals (Bedau [1991, 654]).

He goes on to suggest that the reason this claim seems counterintuitive is that clay crystals are not *living*. The reason that functions are typically only ascribed to the parts of living things, he continues, is because living things can be said to possess a *good*, namely the continuation and propagation of their own existence and that of their kind (Bedau [1992, 801]; see also Fulford [1999, 416-7], who also appeals to non-biological examples to undermine a non-evaluative notion of function, and McLaughlin [2001, 181-2], who also uses the crystal example for a similar end). What this intuition reveals, according to Bedau, is that all function assignments possess an implicit evaluative element – that is, the function of an entity consists, in part, in its present or past contribution to some *good*.

But this conclusion would not bode well for the attempt to construct a theory of function that satisfies CA_1 and CA_2*. The reason is that, if the notion of function is *evaluative* in Hare's sense (as described in Section 1.3), then any intelligible function ascription presupposes an act of commendation on the part of the person who utters it. In other words, to say that the function of *X* is *Y* is to say that *X*'s doing *Y* has a *good*

consequence – typically associated with survival or reproduction – and hence to endorse a set of values concerning the nature of the good, or the nature of individual well-being. Consequently, the warranted ascription of "function" would depend centrally upon the attitude of the speaker; hence, whether or not something can be said to possess a "function" would depend partly upon whether people who assign functions to entities conceive of survival and reproduction as good things.

However, there are two responses one might make to this suggestion. The first is to reject the validity of the counterexample by accepting, in accordance with SRE, that clay crystals *do* possess functions. This is Millikan's (1993) response. As she claims, the clay crystals discussed in Bedau's example certainly possess *functions*; however, they do not possess *biological* functions. As noted above (Section 4.1.3; *fn*. 116), in her view there exists a unified notion of function that applies to artifacts, behaviors, social institutions, and biological structures. She argues that the counterintuitive appearance of Bedau's example stems from the fact that he is mistakenly attributing to the parts of the crystal *biological* functions – which, by definition, they cannot possess. However, there is no reason it cannot possess a more generic type of function: "….if crystals can have functions, as well as learned behaviors, artifacts, words, customs, etc., that is fine by me" (Ibid.; see *fn*. 7 of that text).[126] Sarkar (2005) offers a similar response, claiming as a virtue of his theory of function (see Section 3.1.2, *fn*. 74) that it can apply to inanimate entities. In his view, this should be seen as a welcome result, given that "there is no principled distinction between living and nonliving matter" (Ibid., 40; see *fn*. 57 of that text).

[126] Millikan also argues that it does not seem counterintuitive to ascribe functions to viruses, even though they do not necessarily qualify as "living"; consequently it does not seem to be the case that functions are only ascribed to living entities (Ibid.).

A second response is to accept the validity of the counterexample, and hence that a non-evaluative, purely etiological theory is insufficient for defining "function". However, even if one does not accept that the etiological theory is sufficient for defining "function", one can argue that it is *necessary* for defining "function". This argument would proceed in two steps. The first step of the argument is to assume that the concept of function possesses an evaluative element, but to argue that this evaluative element is *also*, by itself, insufficient for defining "function". As argued in Section 3.1.3, under "Good-Contribution Theories", the view that the function of an entity consists in that activity that contributes to some systemic *good* cannot alone distinguish the function of an item from a fortuitous benefit that it produces. The function of the nose is not to hold up glasses, even if holding up glasses represents a good for the individual. The second step is to argue that the insufficiency of a purely evaluative theory of function can or should be resolved by incorporating an etiological element into one's definition of function. For example, one would argue that the reason that the nose does not have the function of holding up glasses is because, even though holding up glasses is a good consequence of possessing a nose, the capacity of the nose to hold up glasses does not explain the current presence of noses. As noted above, Ayala (1970), Bedau (1992, 799), and McLaughlin (2001, 168) all present such a "mixed" or hybrid etiological-evaluative view according to which the function of an entity consists in that activity that, in the past, contributed to some systemic good (e.g., survival or reproduction) and, as a consequence, explains the (differential) reproduction or persistence of that entity or type of entity.

The standpoint of this dissertation is that the notions of function and dysfunction are not evaluative, and that Bedau's counterexample is not alone a persuasive reason for accepting an evaluative notion of function. However, even if one finds his counterexample persuasive, as long as one accepts that any adequate theory of function

must possess an etiological component – at least as a necessary condition – then the central claim of this dissertation remains unaffected, for the following reason. According to this dissertation, the available evidence concerning the neurobiological basis of schizophrenia provides little reason to suppose that this etiological condition on the concept of function will be fulfilled. As a consequence, the claim that schizophrenia stems from a dysfunction on the part of the brain is unwarranted. Therefore, even if one wishes to supplement the concept of function by introducing an evaluative element, in addition to the etiological element, the claim that schizophrenia stems from a dysfunction on the part of the brain will remain unwarranted. Consequently, for the purpose of the dissertation, it ultimately may not matter whether one accepts an evaluative or non-evaluative definition of "dysfunction", as long as one accepts that an etiological component is a necessary condition for such a definition.[127]

4.1.5 Adequacy of SPE on All Three Grounds

According to SPE, in order for something to have a function, it must have contributed to the differential persistence or reproduction of that entity or type of entity. With few exceptions, the logical possibility of SPE has been ignored in discussions of etiological theories of function. This is most likely due to the fact that natural selection is often thought to operate exclusively over *reproducing* entities, and hence that any strong theory of function must be identical to SRE. However, as pointed out in the previous section, synaptic structures, as well as some learned dispositions, are partly formed by

[127] Wakefield (1999b, 470-1) essentially makes this same argument in response to Fulford's (1999) claim that the notion of a consequence-etiology is not *sufficient* for defining function (Ibid., 416), and that some prescriptive element is necessary to that definition (Ibid., 417). Wakefield argues that even if some prescriptive component is a necessary condition on the analysis of "function" – and hence that function is an evaluative concept – so long as some reference to a selection process is *also* a necessary condition on this analysis, then the implications of the analysis are largely unaffected with respect to the sort of *biological* evidence that would be required to warrant the inference that a trait is functional (or dysfunctional).

selection processes that operate over non-reproducing entities, and that bring about the *differential persistence* of those entities.

There are three advantages associated with accepting SPE. First, SPE resolves the Boorse-style counterexamples. As noted earlier (Section 4.1.2, under "Intuitive Implausibility of WPE"), none of Boorse's counterexamples illustrate a selection process. For example, although obesity contributes to a sedentary lifestyle and thereby to its own persistence, obesity was not selected over some other trait because it contributed to a sedentary lifestyle. Similarly, the leak in the hose is not there because *it*, rather than something else, proved more effective in knocking out scientists. Because SPE excludes Boorse's counterexamples, it is more intuitively plausible than WPE. Second, SPE is more general than SRE because it does not contain as many unmotivated restrictions. Hence, it is consistent with the methodological bias toward making the fewest number of potentially questionable assumptions. Third, and most importantly, it permits functions to be assigned to unique, non-reproducing structures, and hence is more compatible with the ascription of functions in the context of psychology and neuroscience. (Section 4.2 will be devoted to demonstrating the empirical plausibility of this claim.) Hence it better satisfies the pragmatic motivation of the dissertation.

The remainder of this section will show that SPE, or theories of function very similar to it, has played an important role in philosophical understanding of the relation between teleology and learning. SPE was first formulated in its full generality by Wimsatt (1972), but it also plays a role in some earlier and later attempts to characterize the purposeful character of learning and synaptic structure. Thus, the purpose of this section is to establish the philosophical precedents for SPE.

Wimsatt's (1972) Formulation of SPE

The first explicit and comprehensive statement of SPE is Wimsatt (1972), although this theory is foreshadowed in earlier philosophical approaches to teleology (see below). In that paper, he makes the claim – one that is familiar by now – that natural selection can ground teleological explanations for the current existence of certain traits:

> The class of functions of biological adaptations comprises exactly those functions and *some* of those adaptations…whose existence, presence, and form are purportedly explained (in some sense) by evolutionary theory through the operation of natural selection…*Explanations of the above type are teleological explanations*. (Ibid., 7-8)

However, Wimsatt is clear that "selection" does not only operate over reproducing entities; it also serves to bring about the differential reinforcement or persistence of non-reproducing entities within a system. His paradigm example is learning. At the most abstract level, he claims, all learning can be modeled as the outcome of selection processes:[128]

> In each case [of learning] there are two correlative processes involved in selection, aptly named 'blind variation' and 'selective retention' by Campbell…The concept of progress through trial and error is virtually synonymous with 'blind variation and selective retention', as progress is presumably the result of those two processes as reflected in what is selectively retained as the agent or system progresses through a number of trials…[T]here is good reason to say that *all* problem-solving behaviour has a basic trial and error character. (Ibid, 14)

According to Wimsatt, then, what makes a learned disposition "purposeful" is the same as what makes a biological adaptation "purposeful", namely, that the current

[128] See below, which argues that this claim is implausible: there are types of learning which do not appear to involve any process analogous to natural selection.

existence of the structure is partly explained by a selection process, rather than by any of the consequences that the structure currently produces: "[T]he operation of selection processes is not only *not* special to biology, but appears to be at the core of teleology and purposeful activity wherever they occur" (Ibid., 13).

The idea that learning is a teleological process, that is, the view that what makes a learned disposition *purposeful* is the historical process that produced it, rather than the future consequences it brings about, had been proposed earlier by Mace (1949 [1935]) and Scheffler (1966 [1958]). It has also been proposed in some recent approaches to teleology, in particular, by Godfrey-Smith (1992) and Papineau (1994). Although all four of these philosophers either implicitly adopt SPE, or some theory of function very similar to it, none of them succeed in grasping SPE in its full generality, or in clearly separating WPE and SPE in the context of learning.[129] By presenting their views, then, one can more clearly delineate the precise content of SPE theories.

Teleology and Learning

Mace (1949 [1935]) holds that learned behaviors constitute paradigmatic teleological phenomena, and claims to show that their purposeful character can be defined without appealing to the presence of conscious intentions. Since Mace defines a "teleological system" as one that is "constructed by a teleological process" (Ibid., 535), his analysis is an etiological one – it involves an appeal to the *history* of the system. Clearly, the weight of his analysis rests upon his explication of "teleological process".

To illustrate a teleological process, Mace refers to an ethological example that can be characterized by Thorndike's "law of effect" (Thorndike [1911, 244]). According to

[129] As Buller (1998) has pointed out, many philosophers have vacillated between strong and weak etiological theories; this is all the more tempting when the strong and weak etiological theories are applied to learning, since the distinction between the mere *reinforcement* of a learned disposition and the *differential reinforcement* of that disposition is not as obvious as the difference between the reproduction of a trait and its differential reproduction.

this law, in a given situation, those behavioral responses that are followed by a reward (or "satisfaction") will be more likely to recur in that type of situation, those followed by "discomfort" will be less likely to recur, and the likelihood of recurrence will be proportional to the intensity of the reward (or "discomfort"). The analogy between the law of effect and operant (or instrumental) conditioning is clear, with "satisfaction" and "discomfort" replaced by "positive" and "negative" reinforcement. Like operant conditioning, the law of effect characterizes learning as a selection process in which behavioral patterns from a pre-established repertoire are differentially reinforced by virtue of their relative performance on some common criterion.[130]

Mace's definition of "teleological process" simply generalizes the form of this process. What is crucial to a teleological process, Mace argues, is the presence of a condition *E*, and a set of possible actions, such that those actions which increase the probability of *E* tend to be continued or repeated, and those that decrease *E* discontinued, as a result of which the former are "stabilized" and the latter "eliminated" (Ibid, 535-6):[131] "Any train of actions conforming to this description would, I think, be commonly described as a conative or teleological process" (Ibid., 536). Hence, by appealing to a selection process to define "teleological process", Mace implicitly accepts a version of SPE. Interestingly, Mace does not extend this generalization to cover other biological phenomena, such as evolution by natural selection, that would also fall under his schema. Hence, unlike Wimsatt (1972), Mace does not recognize the generality of his definition.

---

[130] In fact, B. F. Skinner, in his seminal *Science and Human Behavior* (1953), notes the analogy between natural selection and operant conditioning: "We have seen that in certain respects operant reinforcement resembles the natural selection of evolutionary theory. Just as genetic characteristics which arise as mutations are selected or discarded by their consequences, so novel forms of behavior are selected or discarded through reinforcement" (Ibid., 430). (Also see Skinner [1981], for a short article devoted to an elaboration of this analogy.)

[131] Mace introduces other conditions that, from the point of view of this dissertation, are extraneous.

Scheffler's article (1966 [1958]) is primarily written for the purpose of criticizing existing goal-contribution theories of teleology, specifically, those of Rosenbleuth *et al*. (1943) and Braithwaite (1953) (see Section 3.1.3., under "Goal-Contribution Theories"). His own etiological analysis stems from that criticism. According to the "cybernetic" account of teleology, proposed in Rosenbleuth *et al*. (1943), a behavior qualifies as purposeful if it is governed, in part, by a negative feedback process that ensures the attainment or maintenance of some goal-state.[132] For example, a homing torpedo is controlled, in part, by acoustic signals from the object that it eventually destroys. A problem with such theories, Scheffler points out, is the "problem of the missing goal-object"; that is, there are instances of purposeful behavior that are not, in fact, controlled by the putative "goal-object" (Scheffler [1966 (1958), 52]). For example, if an infant cries for his or her absent mother, this behavior is clearly purposeful, but it is not governed by any signals from the "goal-object", namely, the mother. Similarly, if a rat depresses a lever in order to obtain food, but the food box is empty, then the behavior is still teleological even though it cannot be controlled by signals emitted from the non-existent food. According to Scheffler, the problem of the missing goal-object suggests that what makes the behavior purposeful is not that it is *presently* controlled by a feedback mechanism but that, *in the past*, the behavior in question was associated with some reward and thereby came to be reinforced, and that this history of reinforcement explains its current manifestation. The reason the infant's crying qualifies as a purposeful act, he states, is that:

> Having initially cried as a result of internal conditions *C*, and having thereby succeeded in attaining motherly solace, representing a type of rewarding effect *E*,

[132] In the following, the "cybernetic" account of teleology will refer to that version of the goal-contribution account that explicitly refers to the operation of a negative feedback process (see Section 3.1.3; under "Goal-Contribution Theories").

> the infant now cries in the absence of *C*, and as a result of several past learning sequences of *C* followed by *E*. The infant's crying has thus been divorced from its original conditions through the operation of certain of its past effects. (Ibid., 53)

Moreover, he points out that such an account of purposefulness allows teleological explanation to be "explanatory" – in the causal sense, where the explanans temporally precedes the explanandum – while permitting an apparent reference to a future effect of the behavior, thus resolving the problem of backwards causation:

> The apparent future-reference of a teleological description of this present interval is thus not to be confused with prediction...Rather, the teleological statement tells us something of the genesis of the present crying...Such an account is perfectly compatible with normal causal explanation. (Ibid.)

Thus, Scheffler's rationale for proposing an etiological account of teleology in the context of learning is identical to Ayala's (1968; 1970) and Wimsatt's (1972) later rationale for endorsing an etiological account of function in the evolutionary context, namely, that it resolves the problem of backwards causation by introducing a consequence-etiology.

Strictly speaking, however, since there is no reference to *selection* in Scheffler's account, it qualifies as a version of WPE, rather than SPE. For example, the fact that the infant's *crying* was reinforced because, in the past, crying attracted its mother, does not imply that there existed a set of variant behavior patterns culled from a pre-established repertoire (e.g., grasping, making sucking motions, etc.) that were discontinued because they failed to attract its mother. Just as an entity can reproduce without undergoing differential reproduction, something can be reinforced without being *differentially* reinforced. However, if one amends Scheffler's suggestion to require that the infant's

crying must have been reinforced over some variant behavior, one would have an SPE view.

Perhaps the only recent attempts to characterize learning as a teleological process are those of Godfrey-Smith (1992) and Papineau (1994). (This exposition will first describe Papineau's view because it is more developed.) Papineau begins his argument by adopting the premise that natural selection can be thought of as a source of "design" (Ibid., 77) – though one that clearly does not require the existence of a "designer". He then argues that individual *learning* should also be thought of as a source of "design", because it, too, involves a form of selection. The neuronal mechanisms that mediate learned dispositions, he claims, have been *selected for* because of their capacity to produce the disposition in question: "I take it also that this neuronal mechanism was selected (reinforced, developed) *because* it produced that [disposition]" (Ibid., 78). Thus, Papineau echoes Wimsatt's (1972) reasoning that it is the *selective* character of learning that underlies its ability to produce purposeful structures.

Godfrey-Smith (1992) endorses a similar viewpoint, and argues that it is the *selective* character of learning which allows one to assign functions to learned dispositions: "It is important that the selective approach [to defining "function"] is in no way tied to the genetic kind of biological evolution…A selective basis for functional characterization is available whenever learned characters are maintained within the cognitive system because of their consequences" (1992, 292). Both Godfrey-Smith and Papineau, then, clearly *state* that they recognize a formal analogy between natural selection and learning in the context of teleology. However, neither fully draws out the consequences of this analogy. This is because neither of them recognizes a distinction between the claim that a given neural connection is "reinforced", "developed", or "maintained" by virtue of one of its consequences, and the claim that the connection is

selected for – that is, reinforced *over* some other connection. (Section 4.2.2 below will clearly illustrate the difference between these two scenarios in the neurobiological context.) Hence, neither clearly distinguishes between SPE and WPE.[133]

The shortcomings with Papineau's (1994) and Godfrey-Smith's (1992) attempts to tie together natural selection and learning suggest the need for formulating more explicitly what is required of a process in order for it to qualify as a "selection process". This will be particularly crucial for sifting through the range of theories – both in psychology and neurobiology – that purport to be "selection" theories. As will be shown in Section 4.2.1, and as opposed to Wimsatt's assessment, not all learning can be modeled as a selection process. Similarly, as will be shown in Section 4.2.2, and as opposed to a remark made by Crick (1989) that "almost everybody's theory could be called a theory of synaptic selection" (Ibid., 247), not all processes of synapse structure formation are selection processes either. An evaluation of selection processes in learning and neurobiology will be important, then, for assessing the nature and scope of selection in these fields and hence the applicability of SPE as a useful notion of function for these fields.

4.2 SELECTION PROCESSES IN PSYCHOLOGY AND NEUROBIOLOGY

This section will present a brief overview and critical assessment of two claims. The first claim is that learning involves a selection process (Section 4.2.1); the second is

[133] The fact that Godfrey-Smith (1992) does not clearly distinguish between WPE and SPE becomes evident in later articles that are devoted more specifically to teleology in the evolutionary context. After noting the Boorse-style counterexamples that afflict WPE, he rejects WPE in favor of WRE, thus ignoring the logical possibility of SPE. Although he deems this move to WRE necessary for avoiding those counterexamples, he notes that it is unfortunate, because, by restricting "functions" to entities that are capable of reproduction, it breaks the connection between teleology in the context of learning and neuroscience, on the one hand, and evolutionary biology, on the other. For example, he points out that it entails the rejection of Dretske's (1988) theory of function, which is explicitly formulated to be consistent with usage in the neuroscientific context. The fact that the distinction between SPE and WPE was not clearly made, however, prevented him from observing that WRE is not necessary for resolving Boorse's counterexamples.

that synapse formation involves a selection process (Section 4.2.2). This exploration will be useful for evaluating the conditions under which the claim that a given structure or disposition has undergone selection – and hence that it has a function – can be empirically tested. Two conclusions will be drawn from the considerations raised in this section. The first is that, although evaluating the empirical claim that a given neural structure or learned disposition has undergone selection may be technically difficult, there is no principled difficulty of the sort that affects the attempt to reconstruct the long-term evolutionary history of a trait. Hence, SPE avoids the insurmountable empirical problems that afflict SRE (see Section 4.1.3). The second conclusion that will be drawn is that one cannot infer from certain *maladaptive* consequences of a psychological or behavioral pattern or process that it stems from an internal *dysfunction*. Since some of these patterns or processes may be produced by learned responses or neural structures, they may represent the outcome of a selection process, and hence, from the point of view of SPE, they may represent instances of proper *functioning* rather than malfunctioning. This consideration raises the level of evidence that would be required in order to warrant the claim that a given psychological or behavioral condition stems from an internal dysfunction.

4.2.1 Learning as a Selection Process

The observation that underlies Wimsatt's (1972) version of SPE is that there exists an analogy between learning and natural selection. That *some* forms of learning are analogous to selection processes is not at issue – as noted in the previous section (Section 4.1.5, under "Teleology and Learning"), the differential reinforcement (amplification or extinction) of learned dispositions brought about by positive and negative reinforcement exhibits all of the characteristics of a selection process. Hence, if there is a problem with characterizing learning as a selection process, it is not that examples will fail to be found.

Rather, it is that *all* learning may appear to have a selective character, and this would trivialize the substantive content of the analogy.

An example of the way in which the notion of selection can be overgeneralized to include all forms of learning can be seen in Herbert Simon's (1969) *The Sciences of the Artificial*. Simon argues that, in its most general form, all human problem solving consists of a mixture of trial-and-error exploration and some form of feedback. Moreover, he explicitly draws the analogy to natural selection:

> Human problem solving, from the most blundering to the most insightful, involves nothing more than varying mixtures of trial and error and selectivity. We do not need to postulate processes more sophisticated than those involved in organic evolution to explain how enormous problem mazes are cut down to quite reasonable size. (Ibid., 99)

Wimsatt (1972) was clearly influenced by this text when he wrote that "[T]here is good reason to say that *all* problem-solving behaviour has a basic trial and error character" (Wimsatt [1972, 14]).

The psychologist Donald Campbell – one of the founders of evolutionary epistemology – has taken this selectionist perspective to extremes, thereby illustrating the risk of trivializing the analogy. Throughout several articles that span many decades (e.g., Campbell [1956; 1960; 1974; 1988]), Campbell argues that the two-fold process of "blind variation and selective retention" is at work at all levels of biological, psychological, and social development – for example, in locomotion, perception, imitative behavior, language, and even the process of scientific theorizing. In summarizing his view, he writes, "Human knowledge processes, when examined in continuity with the evolutionary sequence, turn out to involve numerous mechanisms at various levels of substitute functioning, hierarchically related, and with some form of selective retention process at

each level" (Campbell [1974, 419]).[134] Cziko (1995), similarly, develops a "universal selection theory" that postulates the operation and efficacy of selection processes at all levels of biological and social organization.

Nonetheless, however instructive the analogy may be, it seems to become more tenuous as one moves up the scale of psychological and sociocultural development. For example, is the "natural selection of scientific theories" a metaphor or a fact? Thagard (1988, 101-11), for example, criticizes the analogy between theory selection and natural selection on three grounds. First, the generation of variation (new theories) is not "blind", since new theories are usually formulated for the purpose of solving a specific problem (Ibid., 106).[135] Secondly, unlike natural selection in the biological realm, where criteria for differential reproduction are always susceptible to change as a function of the changing environment, the differential survival of scientific theories is partly based on "global" criteria, that is, criteria that hold sway throughout the scientific community, such as simplicity, predictive or explanatory power, theoretical fruitfulness, and so on. This allows a concept of "progress" to be defined for scientific theories that is undefined in the biological realm (Ibid., 108). Finally, there is no obvious parallel to genetic transmission. In science, "preservation is by publication and pedagogy, not by any process resembling inheritance" (Ibid., 109).[136]

---

[134] Campbell attributes the specific analogy between the process of scientific theorizing – in which "unsuccessful" hypotheses are discarded and "successful" ones propagated – and biological evolution through natural selection to Karl Popper. As Popper writes, "The method of trial and error is not, of course, simply identical with the scientific or critical approach – with the method of conjecture and refutation. The method of trial and error is applied not only by Einstein but, in a more dogmatic fashion, by the amoeba also" (Quoted in Campbell [1974, 416]).

[135] Also see Amundson (1989, 427) for a similar criticism.

[136] Hull (1988), however, defends the analogy between the dissemination of a scientific theory and genetic transmission. In the biological case, for example, genes may serve as "replicators" and organisms, "interactors"; in the scientific case, theories and methods are the "replicators" and the scientists themselves, the "interactors" (Ibid., 139-143). It is the scientists rather than the theories, therefore, that "compete for success" in Hull's view. He also argues that although the intentional and progressive character of science makes it disanalogous with biological evolution, it does not for that reason make it non-selectionist (Ibid., 145-147).

Consequently, in order to ensure the usefulness of the concept of selection in the context of learning, it is necessary to show that the concept is, in fact, discriminating – that is, that it can only be applied to certain learning processes and not others. Perhaps the best way to show its usefulness, then, is to reveal its limitations. Although this section will not delve very deeply into different theories of learning, it will at least set up a very general contrast between theories of learning that involve selection and those that do not, and hence preserve some non-trivial and substantive content to the claim that learning can be a selection process.

As noted earlier, the most obvious type of learning that can be modeled as a selection process is learning that is mediated by positive and negative reinforcement. Skinner (1981) addresses this analogy between natural selection in the context of biological evolution and operant conditioning, and claims that the pivot of the analogy is that both processes involve "selection by consequences"; that is, in both processes, the consequences of a disposition affect its chances of being differentially retained or perpetuated. But clearly, not all forms of learning can be modeled in this way! The most obvious and uncontroversial example of a non-selective learning process is classical conditioning. In classical conditioning, the "conditioned stimulus" (e.g., a tone) is paired with an unconditioned stimulus (e.g., food); by being presented to a subject simultaneously or in short succession, the conditioned stimulus comes to elicit the same response (e.g., salivation) as the unconditioned stimulus. However, salivating in response to a tone is in no sense "selected over" some other response (e.g., barking) because it, rather than the other, was more effective at procuring food. This is because the item that qualifies as a reinforcer in this example (food) is not withheld in the absence of salivation, and hence even if the subject were to have produced varying behaviors, there is nothing analogous in this model to the differential fitness of a behavioral response.

Observational learning (or imitative learning) provides another example of a learning process that does not necessarily involve selection. In observational learning, a disposition can be acquired through the observation of a model and the retention of certain details of the modeled behavior. In this situation, learning occurs even if the modeled behavior is not immediately reproduced by the learner. But if the behavior is learned before it is reproduced, then this form of learning does not involve reinforcement by the consequences of the learner's behavior. As Bandura *et al*. (1961) point out, "Unless [behavioral] responses are emitted…they cannot be influenced…Indeed, social imitation may hasten or short-cut the acquisition of new behaviors without the necessity of reinforcing successive approximations…" (Ibid., 580).[137]

Despite its limitations, the process of "blind variation and selective retention", abstractly understood, clearly does play a role in certain aspects of learning and therefore possesses a legitimate sphere of application in that realm. Although operant conditioning is not successful as a complete account of human cognitive development, nonetheless, the role of positive and negative reinforcement cannot be excluded from any such complete account, especially as the mechanisms of positive and negative reinforcement become better characterized on the neurobiological level.[138]

According to SPE, if a disposition is retained by virtue of a selection process, then it comes to possess an etiological function – namely, the function of doing whatever it did that led to its differential reinforcement. But this observation has critical consequences for the practice of psychiatric diagnosis and classification. In particular, it entails that the

[137] However, the same authors raise the possibility that, between the observation of the model and the later reproduction of the behavior, the imitator may reinforce the behavioral pattern *covertly*, by, for example, anticipating the pleasure to be gained from carrying out the imitated behavior (Ibid.). Perhaps one might allow this type of learning to involve a form of "selection", but one in which certain behavioral patterns are reinforced because of their *imagined* consequences rather than their actual consequences. It is not clear whether the concept of selection should be applied to this process.

[138] See, e.g., Kelley and Berridge (2002) for an overview of neurobiological theories of reward and motivation.

fact that the psychological or behavioral consequences of a given disposition may be, in some sense, "maladaptive" in a given environment does not imply that the disposition itself is dysfunctional, or that it stems from an internal dysfunction. Rather, given the learning history that explains how that disposition came to be maintained or reinforced, the consequences may instead represent the normal or proper functioning of the disposition. In the latter case, one may intuitively *not* want to diagnose the presence of a disorder.

A simple example, elaborated by the developmental psychopathologists Richters and Cicchetti (1993), serves to illustrate the point.[139] The diagnostic criteria for conduct disorder (APA [2000, 93-99]) include aggression, deceitfulness, and rule-violation. (Conduct disorder is primarily diagnosed of children or adolescents; antisocial personality disorder is reserved for adults, but many of the diagnostic criteria are shared by both categories.) As Richters and Cicchetti (1993) point out, if a child or adolescent expresses antisocial behavior patterns, such as anger, defiance, and oppositionality, this may suggest the presence of a dysfunction, but it may also implicate a developmental context in which those behaviors and attitudes were differentially reinforced. In light of that context, the behaviors in question may appear "normal" or "functional" for that context:

> [S]ome children might develop antisocial behavior patterns in the absence of internal dysfunctions; their conduct problems instead may be caused entirely by extrinsic, environmental factors. An obvious example of this might be children raised in criminogenic neighborhoods and/or families and those who engage in antisocial, even criminal, actions because those are the behaviors modeled, expected, and/or rewarded by the major influences in their environments. (Ibid., 15)

[139] Although see Wakefield's (1992b, 242) discussion of conduct disorder, and Agich's (1994, 242-44) discussion of antisocial personality disorder. Both discussions make the point that behavioral criteria alone are insufficient to distinguish between disorder and non-disorder.

However, if a child evinces these behaviors as early as preschool, and they do not appear to be "engendered or reinforced by inconsistent or deviant parenting" (Ibid., 18), then one would be much more inclined to diagnose a mental disorder.

The above considerations on the notion of function can provide some theoretical structure to support the intuition that the diagnosis of a mental disorder in the first case (that in which the behavior was reinforced) is not warranted, but that it may be warranted in the second case, even though both children exhibit many of the defining criteria of conduct disorder. In the latter case, according to SPE, the problematic behavior represents the proper or normal functioning of a learned disposition that, unfortunately, may produce maladaptive consequences.

The possibility that the apparently maladaptive consequences of a disposition do not necessarily allow one to infer the presence of an internal dysfunction is not a novel result of the foregoing analysis of "function". In fact, it is implied by the so-called "mismatch" argument raised above (see *fn*. 123, this chapter). According to this argument, some of the characteristic psychological conditions studied in psychopathology, such as anxiety or depression, are simply the inevitable consequences of the normal operation of evolved psychological mechanisms that were selected for in ancestral environments but produce untoward consequences in the modern social context. The foregoing considerations on the nature of learning merely provide further support for this distinction. Hence, they raise the standard of evidence that would have to be marshaled in order to show that a given psychological or behavioral condition stems from an internal dysfunction. However, the example of conduct disorder also illustrates that just because a condition does *not* stem from an internal dysfunction does not mean that it is any less meritorious of social concern, psychological counseling, or corrective

treatment. A similar conclusion will be drawn for the neurobiological context in the following section.

4.2.2 Synaptic Structure Formation as a Selection Process

If selection processes are operative at the neurobiological level, then there exists warrant for assigning etiological functions to that level – that is, to structures that may be unique, non-heritable, and capable of functional reorganization in response to novel environmental demands. As Richters and Hinshaw (1999) point out – though without offering any well-defined theory of function – the vast extent of neuronal plasticity in the developing brain implies that the brain may acquire novel functions during epigenetic development and hence that neuronal functions should be defined with respect to "both the evolutionary history of the species as well as the ontogenetic shaping influences of each individual's experience" (Ibid., 441).

There are three reasons to delve into "neural selection" theories. First, the application of etiological theories of function in the neurobiological context is as yet poorly understood – as noted above, both Papineau (1994) and Godfrey-Smith (1992) attempt to apply etiological theories in this realm, yet without adequately working through the consequences of the position. Second, given the increasing relevance of neurobiological results to the psychiatric context, it becomes imperative, from the perspective of this dissertation, to state precisely the conditions under which a neural structure can be said to be "dysfunctional". This requires stating explicitly, and in an empirically testable manner, the conditions under which it comes to possess a "function". Third, in the neurobiological context, "learning" is often used very broadly to refer to more or less permanent, activity-dependent changes in synaptic structure. Hence, in some sense, this section does not represent a departure from the foregoing considerations on the

nature of learning, but an extension of those considerations into the neurobiological context.

This section will begin by explicating the difference between selective and non-selective (e.g., "constructive") types of synaptic structure formation. It will then establish the empirical warrant for considering neural selectionism to be a widely-applicable through not universal theory of synaptic structure formation. Although there is much debate concerning the ubiquity of neural selection, there is no doubt about its existence or significance as an important mechanism of activity-dependent synaptic structure formation.

Selection and Construction

There are two types of processes that account for the mature synaptic structure of the brain and nervous system: "activity-independent" and "activity-dependent" processes. To say that a neuron's pattern of connectivity – that is, its pattern of divergence (the set of neurons it innervates) and its pattern of convergence (the set of neurons that innervate it) – is "activity-*independent*" is to say that this pattern is not based on the activation of that neuron – i.e., the production of electrical potentials and the release of neurotransmitter by that neuron. To say that the pattern is "activity-*dependent*" is to say that the pattern is based, in part, on the frequency and pattern of neural activation itself. "Neural selection" and "neural construction" refer to two types of activity-dependent processes.

An example of a theory according to which synaptic structure formation is largely due to activity-independent processes is the chemoaffinity hypothesis associated with the work of Roger Sperry. According to the chemoaffinity hypothesis, each neuron possesses a specific chemical "marker", established genetically or in the early stages of neural development, and that neuron is guided to a specific target that bears an identical or

complementary such marker (e.g., Sperry [1951; 1963]; also see Meyer [1998]). This theory was originally formulated on the basis of research, described in the previous section (see Section 3.2.1, under "Substantive Approaches Violate Adequate Conditions"), on the effects of the ablation of retinotectal connections in the frog. As is well-known, in the mature frog, the original topographic mapping is eventually reconstituted despite gross alterations in the position or structure of the retina (Sperry [1944]). This suggests that each retinal ganglion neuron bears the chemical trace of its original position on the tectum and uses this trace to re-innervate that position. This research had the broader implication that the retinotectal pattern of connectivity is largely invariant and not subject to extensive modification by experience. (See, however, Meyer and Sperry [1976, 113], for an acknowledgement of the role of activity in bringing about the competitive elimination of certain synapses and hence "fine-tuning" the pattern of connectivity.)

A second type of theory that holds synaptic structure formation to be an activity-independent process represents a variation on the strict chemoaffinity hypothesis. It holds that this chemical signal or "marker" is not unique to a given neuron but to a class of neurons, and that the affinity between an innervating neuron and its target exhibits gradation in strength (Meyer and Sperry [1976]). This latter theory helps to account for the ubiquitous plasticity found in the developing brain, which is difficult to account for on the strict chemoaffinity hypothesis (see, e.g., Gaze *et al*. [1963; 1965] for some of the initial research on the retinotectal projection in *Xenopus laevis* embryos that disconfirmed the strong chemoaffinity hypothesis; also Gaze and Sharma [1970] on similar work with goldfish; see Gaze [1974] for a succinct review). A third type of theory that characterizes synaptic structure formation as an activity-independent process refers to the role of purely

mechanical constraints, such as substrate guidance, for explaining the initial pattern of connectivity (e.g., Scholes [1979]).

Clearly, activity-independent neural processes do not generally lend themselves to neural selection, since, assuming that each neuron is able to locate its target in an accurate manner, such processes do not generate substantial variation in the pattern of connections over which selection could then act. However, selection could perform some role in the differential strengthening and weakening of synapses. For example, if several different neurons converge on a given target neuron, the chemoaffinity theory may allow certain synapses to be strengthened, and others weakened, owing to differences in their activity. Moreover, as Sperry notes (see above), if these activity-independent processes are "error-prone", then that leaves some scope for selection processes to "fine-tune" the initial pattern of connectivity.

Activity-dependent processes are those in which the mature pattern of connectivity is partly due to the activation of the neurons involved. As noted above, in the neurobiological context, "learning" is often used very broadly to refer to such activity-dependent changes in synaptic structure. Learning, in this sense, involves a relation of dependence or interaction between neural activity and neural structure: activity partly determines structure. But how, precisely, does neural activity determine structure? Two very general views have emerged that attempt to answer this question, "neural selectionism" and "neural constructivism".

Neural selectionism construes synaptic structure formation in terms of a two-stage, iterated process (Changeux and Danchin [1976]; Changeux [1985]; Edelman [1978; 1987]; Gazzaniga [1992]).[140] The first stage corresponds to the activity-

[140] It is more accurate to denote the theory described immediately below as "synaptic selectionism", because the unit of selection is taken to be the synapse itself, rather than the entire neuron, or even groups of neurons. These latter two types of selection, however, will be described in more detail in the next sub-section, "Evidence for Neural Selection".

independent formation of synapses (synaptogenesis). This produces an initial pattern of connectivity. This process is, to a large extent, both *random* and *exuberant*. That is, this process creates a large repertoire of synaptic variation, much of it non-adaptive. The second stage corresponds to the reduction of variation via the competitive elimination of certain synapses.[141] This latter stage is an activity-dependent process. Certainly, selectionists accept the fact that throughout life, neurons are capable of branching and extending new axons or dendrites (projections). Synaptogenesis is not entirely arrested after the first stage. However, they consider the further branching and growth of new projections, after this initial round of competitive elimination, to represent a re-iteration of the first stage of activity-independent, random, and exuberant growth.

The analogy between natural selection and neural selection is fairly strong. It rests upon an analogy between a population of neurons that innervate the same target neuron, and a population of reproducing organisms in a given environment. The first stage corresponds to "blind variation", and the second, differential fitness. Presumably, in the second stage, the differential retention of projections on a given target is a consequence of the fact that the variant projections possess a differential capacity to access and utilize some common resource that is necessary for their continued retention on the target neuron, such as a trophic substance supplied by the target neuron, or merely physical sites upon which to form synapses (these hypotheses will be explored below).[142]

---

[141] More accurately, the differential retention of certain synapses over others is due to their differential capacity to utilize some common resource. As pointed out by Lewontin (1970, 1) not all natural selection can be modeled as a type of "competition". For example, if there are two strains of bacteria in a test tube that have different rates of reproduction, then one strain will become more frequent, and the other less so, even if there is nothing corresponding to a limiting resource over which the two must "compete". Although the notion of competitive elimination will be applied throughout subsequent discussion, this qualification should be kept in mind.

[142] Hence neural selection satisfies Lewontin's (1970, 1) criteria for a selection process (also see Darden and Cain [1989, 121-123], who make this connection).

The second main position on how neural activity translates into neural structure is called "constructivism". According to constructivism, synaptogenesis itself is often an activity-dependent, non-selective process (e.g., Purves [1994]; Purves *et al*. [1996]; Quartz and Sejnowski [1997]). Hence, constructivism emphasizes the role that neural activity plays in the *formation* of new synapses, rather than (or in addition to) the elimination of existing synapses. For example, suppose that neuron *A* synapses onto neuron *B*. The activation of *B* by *A* may trigger the growth and extension of new dendrites on *B* and new axon terminals on *A*. This would increase the number of new synapses between *A* and *B* in an activity-dependent manner. Moreover, this is not a selection process, since it involves a mere proliferation of new synapses, rather than the differential proliferation of new synapses. These newly formed neural projections may also branch, extend, and form synapses with neighboring neurons. In this case, the joint activity of *A* and *B* promotes the formation of new synapses in the absence of selection. Hence, constructivists view neural growth in terms of the gradual, progressive, and activity-dependent elaboration of novel synaptic structures and circuitry, rather than the elimination of "excess" circuitry.

This does not imply that constructivists deny altogether the existence of selective elimination of existing synapses. They acknowledge that the constructive formation of new synapses is partly stochastic and "error-prone", and hence will require the selective elimination of useless or maladaptive connections (Purves [1994, 68]; Quartz and Sejnowski [1997, 550]) However, the "directed" quality of synapse formation minimizes the need for a consequent phase of selection.

In short, according to selectionism, synapse formation is profligate and "blind", thus maximizing the need for selection processes to reduce this abundance by preserving those connections that are in some sense "adaptive". It assumes that activity-independent

neural processes have produced the repertoire of variation over which selection will act. A representative quote nicely summarizes this perspective: "To learn is to stabilize preestablished synaptic combinations, and to eliminate the surplus" (Changeux [1985, 249]). According to constructivism, synapse formation is parsimonious, in that it extends and reinforces active neural projections in a way that is sensitive to the functional demands of the system. This minimizes the necessity for selection processes to reduce variability. Purves (1994) gives a fairly clear statement of this viewpoint: "Whatever its cellular and molecular basis turns out to be, activity-dependent growth provides a richer and more consistent framework for thinking about neural development than the now popular idea that we start life with an initial excess of connections and then select from this surfeit by competitive mechanisms akin to natural selection. Rather, the brain builds the circuitry it needs during its progress to maturity" (Ibid., 93-4).

The distinction between neural selectionism and neural constructionism illustrates the distinction between SPE and WPE as alternate etiological theories of function. According to selectionism, neural activity brings about the *differential reinforcement* or retention of synapses through the elimination of specific synapses. According to constructionism, neural activity primarily *extends and reinforces* existing patterns of connectivity through the progressive elaboration of projections. However, since it does not typically involve selection, it explains the reinforcement and development of synapses without appealing to the *differential* reinforcement or development of synapses. The failure to make this distinction is precisely the difficulty alluded to above in relation to Papineau's (1994) and Godfrey-Smith's (1992) attempts to explicate a theory of function that would be appropriate for neural development as well as biological evolution. Hence, if one accepts SPE, then mechanisms of neural construction, since they only involve the uniform strengthening of synapses or the elaboration of existing

projections, cannot bestow novel functions onto the neural structures that it creates. Only the mechanisms of neural selectionism that bring about new patterns of connectivity via the *differential* retention of synapses can qualify as bestowing novel functions onto the structures that it creates.

There exist well-documented cases of both types of processes – selective and constructive – in the brain, and to this extent there is no genuine dispute between constructivists and selectionists with respect to whether either type of mechanism *exists*. In fact, in most cases that will be discussed in the next section, both processes are at work concurrently. The debates, then, involve the relative prominence of one mechanism over the other. Some selectionists, such as Changeux (1985; 1997), Edelman (1978; 1987), Gazzaniga (1992), and Sporns (1997a; 1997b), argue that virtually all synaptic structure formation is selectionist in character, while those such as Purves (1994), Purves *et al*. (1996), and Quartz and Sejnowski (1997), emphasize its constructive character. Katz and Shatz (1996), LeDoux (2002), Black and Greenough (1986; 1997), and Elliott and Shadbolt (1997) emphasize the concurrent operation of both processes and do not argue that one of them is more ubiquitous than the other, but that both are complementary or even inextricable from one another.[143]

Evidence for Neural Selection

There are three different types of "neural selection" depending on what neural structure qualifies as the unit of selection – that is, the unit that is differentially retained

[143] Black and Greenough (1986) propose this "conciliatory" position in their definition of two types of learning, "experience-expectant" and "experience-dependent", where the former refers to the learning of general information available to most members of a species, and the latter to the learning of idiosyncratic information. They suggest that the former type of learning is mainly implemented by neural selection processes involving the activity-independent proliferation and activity-dependent elimination of specific connections, and that the latter is implemented by constructive processes that involve activity-dependent growth and extension of new synapses. Bolhuis (1994, 37-9), however, challenges the claim that these two types of learning can be related in any simple way to specific neural mechanisms.

or reproduced through selective activity. Selective activity may operate over the *synapse* as the unit of selection, the *neuron* itself as the unit of selection, or *groups of neurons* as the unit of selection. Each of these possibilities will be described below, although the main emphasis will be on synaptic selection.

Synaptic Selection

The initial evidence for synaptic selection came from studies of the neuromuscular junction in mammals. At birth, each muscle fiber is typically innervated by more than one motor neuron. Over the course of several weeks, innervating motor neurons retract and a one-to-one pattern of connectivity emerges between motor nerve and muscle fiber (see Purves and Lichtman [1980] for an early review of this research; also see Figure 4.6). In skeletal muscle of newborn rats, for example, each muscle fiber is innervated by two or more neurons; within two weeks after birth each fiber is innervated by only a single neuron (Brown *et al*. [1976]). However, the number of existing motor neurons remains constant during this period. This indicates that the net loss of innervating axons is not due to the death of motor neurons but the retraction or withdrawal of all but one axon (Ibid.). Similarly, submandibular (submaxillary) ganglion cells of newborn rats are innervated by an average of five neurons, and within five weeks the one-to-one pattern of connectivity is established (Lichtman [1977; 1980]).

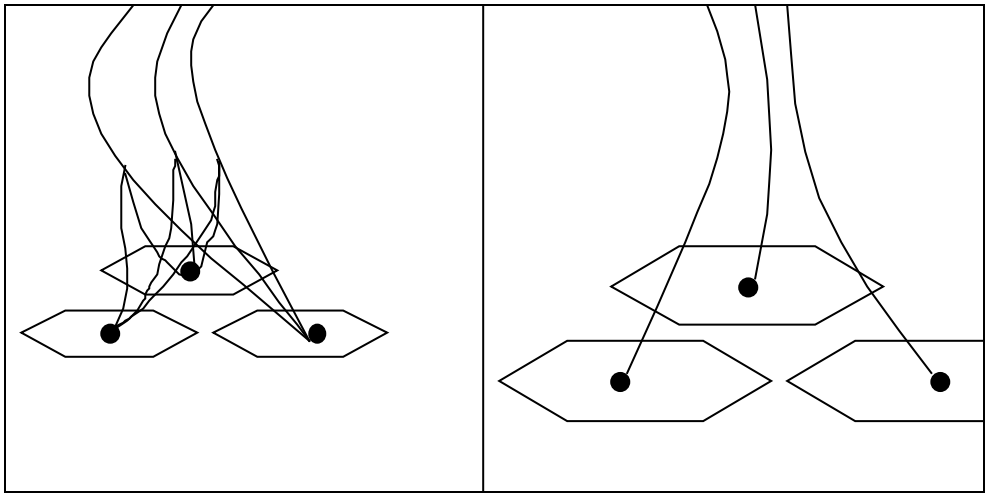

Figure 4.6: Innervation of skeletal muscle of newborn rats. The first panel depicts the multiple innervation of muscle fibers by motor neurons; the second panel depicts the one-to-one pattern of connectivity that emerges by two weeks after birth. Redrawn from Purves and Lichtman (1980, 155).

The authors of the studies referred to above proposed that a competitive mechanism is involved in this eliminative process, and speculated as to the nature of this competition. One such mechanism would involve the production of some trophic material that is synthesized or made available in limited quantities by the target neuron, that is necessary for the maintenance of the innervating axons, and that is taken up by the innervating axons by a retrograde transport mechanism (Brown *et al*. [1976, 420]; Lichtman [1977, 173-4; 1980, 129]; Purves [1977, 38-39]; Purves and Lichtman [1978, 848-851]). (See the next subsection, "Selection of Neurons", on the role of neurotrophins such as nerve growth factor [NGF], or brain-derived neurotrophic factor [BDNF], in preventing naturally occurring neural cell death, and for the possible role of the neurotrophins in synaptic selection). In the case of the neuromuscular junction, any "trait" of the synapse that bestows upon the axon a differential capacity to uptake this trophic substance would give that synapse a competitive advantage in the selection process, and hence, according to SPE, would come to possess the function of doing so.

However, it is important to point out that this eliminative mechanism or selection process is not without its "constructive" and non-selective counterpart in the neuromuscular junction. This corresponding constructive mechanism is attested to by the fact that in rat submandibular ganglion cells, the net *decrease* in the number of innervating axons is followed by an *increase* in the growth and extension of new axon terminals on the remaining axon (Lichtman [1977, 173]; Purves and Lichtman [1978, 830]). These proliferating axon terminals form additional synapses with the target neuron such that there is no net reduction in total synapse number. This wave of synaptogenesis appears to be dependant on neural activation and hence represents a paradigm form of neural constructivism. However, this is a case of neural constructivism that is perfectly compatible with neural selectionism and, in fact, presupposes its operation.

In the central nervous system, the neural process that best illustrates and supports this competitive model of synaptic structure formation involves the development of ocular dominance columns in the primary visual cortex of mammals. In mature mammals, retinal ganglion cells (RGC) from the retina extend axons to the lateral geniculate nucleus (LGN) in the thalamus; these are called "retinogeniculate" projections. LGN neurons, in turn, send axons to layer IV of the primary visual cortex (area 17); these are called "geniculocortical" projections. In some mammals, geniculocortical axons from the LGN form an alternating series of eye-specific "stripes" or "columns" in layer IV called "ocular dominance columns". Although the majority of layer IV neurons respond in some degree to visual stimulation of either eye, cells that occupy the same ocular dominance column respond preferentially to stimulation of the same eye. The formation of these columns appears to fit the two-stage selectionist model that begins with exuberant and random growth, and is followed by the competitive elimination of certain synapses.

This competitive model of ocular dominance formation was first suggested by experiments carried out by David Hubel and Torsten Wiesel in the 1960s (Wiesel and Hubel [1963]; Hubel and Wiesel [1965]). The initial set of experiments suggesting this result simply consisted in depriving newborn kittens of visual stimulation in one eye (monocular deprivation) for the first few months of life, and recording electrical activity from single cells in the visual cortex (Wiesel and Hubel [1963]). In non-deprived newborn kittens, most of the cells in layer IV of the visual cortex are responsive to visual stimulation from either eye – more specifically, about 80% of those cells are "binocularly driven", although a small proportion are exclusively responsive to stimulation of one eye or the other. Hubel and Wiesel then compared normal kittens with monocularly-deprived kittens, and found that in the latter, the vast majority of cells respond exclusively to stimulation from the non-deprived eye alone – they are "monocularly-driven". This comparison suggested that the poverty of connections to the deprived eye came about by the selective loss of connections that were present at birth, rather than the failure of those connections to develop normally in the first place. In other words, normal visual activity appears to "sustain" or "validate" preexisting connections rather than to create them. This is in accordance with an activity-independent model of synaptogenesis.

But how does this relate to the notion that there exists an active "competition" between neurons? Unlike kittens that have undergone monocular deprivation, kittens that have been exclusively dark-reared for the first several months of life appear to retain the same degree of binocularity as normal kittens. This implies that the results of the monocular-deprivation experiments cannot merely be explained by the assumption that connections from the deprived eye degenerate as a function of disuse. Rather, it implies that there exists some active "competition" between the neural connections from the

deprived eye and the non-deprived eye (Ibid., 1015) – that is, the loss of connections from one eye is a function of the activation of the other eye.

These results were confirmed by a second set of experiments, in which Hubel and Wiesel (1965) induced strabismus – nonparallel visual axes – in newborn kittens by severing the extraocular muscle of a single eye. This ensures that the visual signals produced by the two eyes are asynchronous or uncoordinated. The result was that most neurons in the visual cortex responded exclusively to one eye or the other, but not both. Only about 20% of visual neurons were responsive to both eyes, rather than the 80% in normal kittens (Ibid., 1058). This suggested that when confronted with asynchronous patterns of activity, a cortical neuron eventually comes to "favor" the input from one eye over the other: "These results suggest that strabismus caused cells to shift in their ocular dominance, a given cell coming to favor more and more the eye that dominated it at birth, ultimately losing all connections with the nondominant eye" (Ibid., 1058). This result again supported a "competitive" model of ocular dominance formation.

Moreover, they found that the ocular dominance profile of one cell is not completely independent of its neighbor, but that cells with the same profile tend to cluster together in groups (Ibid., 1056), that is, ocular dominance columns. Hubel and Wiesel (1972) later exhibited these ocular dominance columns anatomically in the Macaque monkey by inducing lesions in the LGN that affected only the input from a single eye and then applying a staining method to the visual cortex that selectively stains degenerating geniculocortical axons. The effect of monocular deprivation on the relative width of ocular dominance columns can also be visualized by using autoradiography, through the transport of a radioactive material that has been injected into a single eye (Hubel *et al*. [1977]). Rakic (1976) used this method to demonstrate that, in the fetal Rhesus monkey brain, geniculocortical axons are diffusely distributed and intermixed in the visual cortex,

and that segregation of these axons into ocular dominance columns begins in the second half of gestation. Thus, much of the activity that drives the segregation of ocular dominance columns in the monkey stems from spontaneously-evoked retinal and thalamic activity and not from postnatal visual experience (see Katz and Shatz [1996] for discussion).

The existence of an eliminative process is given further confirmation by visualization of individual geniculocortical axons. As Antonini and Stryker (1993a) show, upon subjecting newborn kittens to monocular deprivation, the retraction and elimination of geniculocortical projections is initiated very rapidly; within 6-7 days after monocular deprivation, geniculocortical axons controlled by the deprived eye are shorter in length and have fewer branches than those controlled by the non-deprived eye. Interestingly, the length and branch number of geniculocortical axons controlled by the non-deprived eye are longer and have more branches than those of a normal kitten (Ibid., 1820). This suggests that ocular dominance formation is not *merely* controlled by a selection process, but by a "constructive" process as well, which involves the activity-dependent branching and growth of new axon terminals as well as the retraction and elimination of others. If there were no activity-dependent growth, then one would expect geniculocortical axons controlled by the non-deprived eye to have the same length and branch number as those of normal kittens. On the basis of their work, Antonini and Stryker (1993b) embrace the conciliatory view that "normal development [of ocular dominance columns] thus appears to involve both selective elimination of widely extended branches and considerable growth and elaboration" (Ibid., 3572).

Thus, in this case as well as in others that will be examined, selection processes and constructive processes should be seen as complementary forms of synaptic structure development. At the very least, the development of ocular dominance columns suggests

that activity-independent growth and selective elimination of *axons* may be followed by the activity-dependent and non-selective proliferation of the *axon terminals* of those axons that have been retained. From the point of view of this dissertation, so long as a selection process plays some role in the development of a given neural system, then one can ascribe functions to features of individual neurons and groups of neurons in that system, even if constructive processes play a significant role in the further development of the system.

All of the examples that have been described so far concern the competition between the *axons* of different neurons that innervate the same target, and they all illustrate and confirm the selectionist hypothesis of an initial, diffuse, and semi-random proliferation of synapses followed by selective synapse loss via axonal retraction. These examples, however, suggest that the *dendrites* of target neurons function as more or less passive recipients of this competitive process. However, the selectionist hypothesis can equally well be applied to the competitive stabilization and elimination of *dendrites* through their differential retraction following an initial, random, and diffuse phase of dendritic branching and synaptogenesis with innervating neurons. (See Wong and Ghosh [2002] for review and discussion.)

For example, the dendrites of retinal ganglion cells (RGC) in mammals, which form the optic nerve that leads from the retina to the lateral geniculate nucleus in the thalamus, undergo an activity-dependent segregation that is similar to that involved in the formation of ocular dominance columns. Initially, RGC dendrites extend diffusely through the inner plexiform layer (IPL) of the retina. Shortly after forming synapses with afferent bipolar cells, RGC dendrites begin to segregate into two distinct sublayers of the IPL, one more proximal and one more distal to the RGC soma. The dendrites that occupy the more proximal layer are innervated exclusively by 'ON' bipolar neurons (those that

are differentially sensitive to the onset of light), and those that occupy the more distal layer are exclusively innervated by 'OFF' neurons (those that are differentially sensitive to the offset of light). This segregation is mediated by glutamate activity, as shown by the fact that glutamate blockage prohibits this stratification and leaves many neurons innervated by both 'ON' and 'OFF' bipolar cells. For example, in cat retinas, after blockage of glutamate activity, about 40% of RGCs remain innervated by both types of bipolar cells, and hence respond equally to the onset and offset of light (Bisti *et al*. [1998]). Wang *et al*. (2001) provides further evidence for the initially diffuse spread and innervation of RGC dendrites in ferret retinas. The segregation of RGC axons in the lateral geniculate nucleus seems to follow a similar pattern of exuberant, activity-independent growth and subsequent elimination (see Lichtman *et al*. [1999, 570-3] for a summary).

A similar selectionist account of dendrite development has been given for the one-to-one pattern of connectivity established between the axon terminals of sensory olfactory receptors and the dendrites of mitral cells in the olfactory bulb of the opossum, *Monodelphus domestica*. Between 5 and 15 days after birth, mitral cell dendrites undergo a diffuse and uniform proliferation toward the dorsal surface of the olfactory bulb, where they form synapses with the bulb-shaped axonal terminals – the "glomeruli" – of olfactory neurons. By day 15, most of the dendrite branches of each mitral cell have retracted, forming a primarily one-to-one pattern of connectivity with the glomeruli (Malun and Brunjes [1996]). Purves (1994), however, describes how the development of the glomeruli themselves, in the mouse olfactory bulb, is characterized by a progressive growth of existing glomeruli and a gradual addition of novel glomeruli, rather than their selective elimination. Unlike the dendrites of mitral cells, nothing in glomeruli development appears to correspond to "an excess of modular circuitry that is

subsequently reduced", but rather, "the persistent growth, and, in some cases addition, of complex neural circuitry" (Ibid., 43). Hence, selectionist and constructionist mechanisms appear to exhibit a type of "division of labor" in the olfactory bulb, where dendrite specification involves selection processes and axonal specification involves constructive processes. This illustrates a case in which selection and construction processes may be compatible with one another.

The formation of dendritic shape, arborization, and stabilization exemplifies not only selective but constructive mechanisms as well. That is, dendrite branching, growth, retention, and retraction involve not only the activity-dependent elimination of "exuberant" growth but the activity-dependent induction of novel growth as well. (See Wong and Ghosh [2002] for review and discussion.) For example, Sin *et al*. (2002) examine the effects of visual stimulation on the growth of dendrites of tectal neurons in the frog. After depriving tadpoles of light for four hours, they exposed them to bright light for four hours. *In vivo* time lapse imaging reveals the growth and extension of new dendritic arbors – thus confirming the constructivist hypothesis of activity-dependent growth – as well as the selective stabilization of existing dendrites. They also show that the dendritic growth is mediated by the neurotransmitter glutamate, and by calcium-dependent signaling mechanisms. This is indicated by the cessation of dendrite growth upon application of antagonists for the post-synaptic NMDA receptor. (Also see Rajan and Cline [1998], who use time lapse imaging to examine dendrite growth and show a role for NMDA glutamatergic activity in dendrite growth.) The influx of calcium through voltage-gated calcium channels (VGCC) also suffices to induce dendrite growth, as shown by the fact that blockade of VGCCs inhibits dendritic branching (Chevaleyre *et al*. [2002]).

Hence, while appreciating the role of selection processes in synaptic structure development, the notion of a universal or generalized neural selectionism must be tempered by an acknowledgement of the extent of activity-dependent growth of dendrites and hence a lessening of the explanatory role of selection mechanisms.

Selection of Neurons

"Selection of neurons" refers to selection processes that operate over the entire neuron itself, that is, those that involve the differential persistence or reproduction of neurons rather than synapses. The existence of one type of selection of neurons appears to take place during neural cell death, which refers to a specific stage of early embryonic development in vertebrates (see, e.g., Cowan [1973; 1978]; Oppenheim [1991]; Johnson and Deckworth [1993]; Pettmann and Henderson [1998]). Neural cell death, or apoptosis[144], appears to involve, in part, a selection process in which neurons "compete" for a limited field of innervation or for a limited number of trophic resources. Some of these resources may act to suppress the genes involved in naturally occurring or "programmed" cell death (Albright *et al*. [2000, S20]).

Neurogenesis in vertebrate embryos involves a phase of neural proliferation and migration followed by a period of widespread cell death. The generality of this phenomenon was first identified by Hamburger and Levi-Montalcini (1949), who observed motor neuron degeneration in the spinal cord of the normal chick embryo. Neural cell death is widespread and occurs in most types of neurons (Cowan [1973], Oppenheim [1991]). For example, up to 40% of the motor neurons in the chick spinal cord undergo cell death (Hamburger [1975]). In the avian nervous system, cell death

[144] The term "apoptosis" was coined by Kerr *et al*. (1972, 241) to distinguish normal or "programmed" cell death from pathological cell death (necrosis) due to, e.g., lesion or infection.

affects 40% to 75% of all neurons, and appears to be distributed evenly across most neural cell types (Cowan [1978, 165]).

The primary function of neural cell death appears to be a quantitative one, which matches the size of a given group of neurons with the size of its innervation field, that is, the population of target neurons or receptors that the group innervates (Cowan [1973]). This is suggested by the long-attested fact that increasing the size of the innervation field through limb transplantation has been shown to increase the number of motor neurons, and decreasing this field through limb extirpation decreases it (Detwiler [1936]; also see Hollyday and Hamburger [1976]). In addition to this quantitative function, it may also serve to eliminate some connections that have been formed by "developmental errors". For example, Clarke and Cowan (1975; 1976) injected a retrograde tracer, horseradish peroxidase (HRP), into the eye of the chick embryo and showed that a small number of neurons in the ipsilateral, rather than contralateral, isthmo-optical nucleus were labeled with HRP. About 80-90% of these labeled neurons die in early development, suggesting that neural cell death performs the qualitative function of eliminating neurons that innervate the "wrong" eye (Clarke and Cowan [1976, 144]). Hughes (1965, 31) also suggests that neural cell death performs this qualitative function.

The quantitative function of neural cell death may be mediated by a "competitive" mechanism. One fairly simple theory is that neurons that successfully innervate a target neuron are preserved, and those that fail to innervate die. Hamburger (1958) suggests such a mechanism when he writes that "The quantitative relationship between the number of motor neurons and the size of the peripheral field of innervation is established by a selective survival of those neurons which find an adequate peripheral milieu, and the degeneration of all others" (Ibid., 399; quoted in Oppenheim [1981, 85]). However, retrograde labeling techniques have shown that neurons that successfully innervate their

target are also subject to cell death (Clarke and Cowan [1975; 1976]). Hence, the selection criterion cannot simply consist in whether or not a neuron innervates a target. This suggests the possibility that, like synaptic selection, a competitive process that involves many different neurons at the site of innervation, and that involves the uptake of a diffusible trophic substance, mediates cell death (Cowan [1973; 1978, 166]). For example, nerve growth factor (NGF), a protein originally isolated from snake venom, was found to contribute to the survival of sympathetic and sensory ganglia in vitro (Cohen and Levi-Montalcini [1956]; Levi-Montalcini and Cohen [1960]), and was later found to occur naturally in rat sympathetic target tissue (Ebendal *et al*. [1980]). Since that time, many different neurotrophins, or secreted factors that promote the survival of neurons, have been identified (Walicke [1989]; Huang and Reichardt [2001]).

Neurotrophins are currently believed to sponsor the survival of neurons not by enhancing cell metabolism, but rather by suppressing a set of genes that are responsible for apoptosis. The evidence for this claim originated from studies on invertebrates. In *C. elegans*, about 10% of somatic cells undergo apoptosis, which is precisely timed and largely independent of cellular interaction (Yuan and Horvitz [1990]). Two genes in particular, *ced-3* and *ced-4*, are specifically relevant to promoting cell death, and a third gene, *ced-9*, appeals to repress or inhibit the activity of *ced-3* and *ced-4*. After the removal of NGF from embryonic rat cells in vitro, neural death was prevented by the addition of inhibitors of macromolecular synthesis, indicating that gene expression is necessary for cell death and hence that neurotrophins play a role in suppressing this gene expression. Similar experiments were carried out in an *in vivo* context with similar results (Oppenheim *et al*. [1990]; see Johnson and Deckworth [1993] for discussion). Since then, two types of specialized synaptic receptors have been found that chiefly interact with neurotrophins (see Purves *et al*. [2004, 553-554], for an overview of these results).

To the extent that neural cell death is mediated by a competitive mechanism, it is not known precisely what variable feature of a neuron gives it a "selective advantage", that is, what trait confers differential survival onto a given neuron. This is determined by what, precisely, the "limiting factor" may turn out to be. On the one hand, according to what might be termed the "production hypothesis" (Oppenheim [1989, 253]), neurotrophins such as NGF are *synthesized* in limiting quantities. Hence, any mechanism that promotes the differential uptake and retrograde transport of neurotrophins may be "selected for", in that it not only increases intracellular availability of those neurotrophins, but also depletes extracellular sources and thereby deprives other neurons of trophic support (Davies *et al*. [1987, 358]; Bothwell [1995, 245]). On the other hand, according to what might be called the "access hypothesis" (Oppenheim [1989, 254]), it is not limited synthesis *per se* which drives competition but limited access to neurotrophins because of a limiting number of available synaptic sites on a target neuron. According to this hypothesis, neural cell death is a result of a competition for space.

Neural cell death is an example of the *differential retention* of neurons; hence, according to SPE, neurons that survive this developmental stage can be said to possess the *function* of doing whatever it was that contributed to their differential retention. Is it also possible to speak of the *differential reproduction* of neurons or neural types, in addition to their differential retention? Although it was once widely held that neurogenesis only occurs during a unique developmental stage and that it does not continue into adulthood, it has been shown recently that neurogenesis does take place throughout life in some areas of the brain, in particular, the hippocampus (for a recent review see Gould *et al*. [1999a]). This raises the possibility that neurons themselves may undergo a selection process involving the differential *reproduction* of neuron types in addition to their differential retention. For example, hippocampal-dependent learning

tasks, such as those involved in spatial orientation, have been shown to preserve newly-generated hippocampal granule cells from degeneration in mature rats (Gould *et al*. [1999a; 1999b]), and hence their survival appears to be dependent in part on their contribution to the learning task. This does not, however, imply that hippocampal granule cells are selected for, that is, that they are reproduced over some other type of cell because they, rather than that other type of cell, contribute to the learning task in question.

Neural Group Selection

A third theory of neural selection that has been proposed is the idea that selection processes operate over large groups of neurons; hence the expression "neuronal group selection" that appears in the subtitle of Edelman's (1987) book-length presentation (Edelman [1978; 1987]; Edelman and Finkel [1984]; Edelman and Tonini [2001]; also see Rosenfield [1986] for a review of several of Edelman's main articles and presentations on the subject). According to this view, one outcome of normal developmental processes is the construction of large repertoires ("primary repertoires") of neural groups, each group consisting of 50 to 10,000 neurons. Each group in the repertoire exhibits a different internal pattern of connectivity but responds in various degrees to the same stimulus pattern (hence they exhibit "degeneracy"). All groups in the repertoire are "isofunctional" because they share a similar response profile, but they are "nonisomorphic" because they differ structurally (Edelman [1978, 64-65]). This set of groups constitutes a primary repertoire over which selection acts. The neural group that responds most specifically to the stimulus pattern that defines the repertoire is differentially strengthened (that is, its intraspecific pattern of connections is strengthened). Presumably, the intraspecific pattern of connections obtaining within

every other group is weakened, or, supposing that each group can belong to more than one repertoire, it takes part in a different selection process defined by another stimulus pattern. The important point is that the environment does not directly induce novel brain structures but selects from among preexisting configurations. As Edelman (1978) writes, "'Selection' implies that after ontogeny and early development, the brain contains cellular configurations that can already respond discriminatingly to outside signals…These signals serve merely to select among preexisting configurations of cells or cell groups in order to create an appropriate response" (Ibid., 54).

As yet, there is little neurobiological evidence for the existence of neural group selection, at least if "neural group selection" is taken to represent a qualitatively different phenomenon than synaptic selection or selection of neurons.[145] Moreover, it is often charged that Edelman, the main proponent of the view, does not as yet adequately explicate the theory itself, despite a large number of semi-popular writings on the subject. This lack of clarity is evidenced by the fact that at least three noteworthy and competent neuroscientists admit their failure to completely comprehend the details of the view (Barlow [1988]; Purves [1988]; Crick [1989])!

Malfunctioning and Maladaptive Neural Circuitry

The conclusion that should be drawn from this section is that synaptic structures, the components of neurons, and perhaps neurons themselves, can have functions, and these functions can emerge as a result of ontogenetic development, that is, by virtue of the processes that explain how the specific structures come to be developed, reinforced, and retained over time. From the perspective of the dissertation, the importance of neural selection processes is that they permit well-defined empirical conditions to be specified

[145] In a more recent presentation of his view, for example, Edelman seems to collapse entirely the distinction between "neural group selection" and synaptic selection, by invoking evidence of the latter to prove the existence of the former (see Edelman and Tonini [2001, 84]).

that can warrant the ascription of functions – and hence dysfunctions – to synaptic structures.

However, as was pointed out in the last section and will be reiterated here, just because a neural structure is in some sense "maladaptive" or unable to perform the function for which it was selected, does not imply that one can attribute to that specific structure an internal dysfunction. For example, as noted above (under "Synaptic Selection"), normal visual experience in mammals brings about the segregation of retinal ganglion cell (RGC) dendrites. One layer of dendrites is innervated by 'ON' bipolar cells, and the other layer by 'OFF' bipolar cells. This allows the visual cortex to respond differently to the onset and offset of light. At birth, however, the neurons are spread diffusely and hence respond equally well to both the onset and offset of light. Now suppose that as a consequence of, e.g., exclusive dark-rearing, the normal visual activity that brings about segregation does not take place. Clearly, such a configuration would be maladaptive in normal environments, since it reduces the individual's discrimination capacity. However, according to SPE, *one should not say of such RGCs that they are dysfunctional even though they fail to perform their species-typical function of responding differentially to the onset or offset of light*. This is because the selection process that bestows this function upon a given neuron has been prevented from taking place, and hence, with respect to discriminating the onset or offset of light, those neurons do not yet possess a function (or, they are unable to perform their function due to an abnormal environment). This implies that from the perspective of SPE, in order to show that a given neural structure is malfunctioning, one must do more than to show that this structure is either maladaptive, atypical, or both maladaptive and atypical. This raises the level of evidence that would be needed to show that a given neural structure is dysfunctional with respect to some capacity.

This chapter brings to a close the "conceptual" part of the dissertation. That is, up until this point the dissertation has largely been concerned with analyzing and selecting a concept of function that would be appropriate to the psychiatric context – though empirical considerations have been relevant to that selection. Equipped with a satisfactory definition of "function" and "dysfunction", the next chapter turns to examples from current biologically-oriented psychiatric research and seeks to answer the following question: does empirical warrant exist for the claim that schizophrenia stems from a biological dysfunction on the part of the brain or nervous system? The argument will be that it does not.

Chapter 5: Schizophrenia and the Dysfunctional Brain

> In some ways schizophrenia can be understood as a disorder of meaning. It is an illness that transforms the commonplace into the supernatural. A radio announcer's mundane chatter turns into cryptic but strikingly personal references aimed at the listener. Newspaper headlines no longer disclose the private failings of politicians and film stars; they communicate the inner life, the biography, of the person with schizophrenia. To the delusional mind, the eyes of strangers in the street confirm conspiracies, challenge secrets, and dispute innocence. Dates in calendars now signal calamities or ratify mythologies. Coincidence is misinterpreted as causation; the irrelevant shouts with significance. What was once trivial is now monumental. The illness appears to construct meaning where there is none and to do so with such virtuosity that psychosis has been mistaken for creativity. Then, after the storm, a strange reversal often takes place: the madness recedes and leaves a profound vacancy of meaning in its wake. It leaves a mind of impoverished thought, diminished action, empty language, and sparse emotion, a mind devoid of imagination and interest. It is as though the very intensity of psychotic experience both generates and then extinguishes all meaning. (Heinrichs [2001, 119])

The previous four chapters were primarily concerned with the conceptual question: what does it mean to say that a given biological process or entity is dysfunctional? More importantly, it asked: in the context of biological approaches to psychiatry, *which* concept of "function" should one invoke when one claims that a given mental disorder, such as schizophrenia, stems from an inner dysfunction on the part of the brain and nervous system? This chapter turns from theory to application, and evaluates whether the claim that schizophrenia stems from an internal dysfunction on the part of the brain or nervous system is warranted. The conclusion will be that it is not. These considerations suggest that the notion that psychiatric disorders, by and large, stem from biological dysfunctions, should be treated with suspicion.

However, this does not mean that is conceptually or physically impossible that schizophrenia can be traced to a neurobiological dysfunction. In fact, this chapter will

provide a precise description of various scenarios that would warrant the claim that a part of the brain is dysfunctional in schizophrenia. However, it will also provide a precise description of scenarios that would warrant the claim that, in fact, nothing in the brain is dysfunctional in schizophrenia, but that the diverse neurobiological abnormalities associated with schizophrenia represent, instead, a brain that is functioning normally or that is unable to function due to abnormal environmental circumstances. The fact that there is currently insufficient evidence to distinguish between the two types of scenarios implies that the claim that schizophrenia stems from a biological dysfunction is currently unwarranted. Hence, this dissertation does not attempt to deny evidence for the existence of neurobiological differences between people with schizophrenia and people without, but rather to reinterpret the significance of those differences.

In contemporary biological approaches to psychiatry it is rarely questioned that schizophrenia stems from a neurological or biological "dysfunction". This dysfunction is often characterized by colorful and imaginative locutions: *The Broken Brain* is the title of a recent book on schizophrenia (Andreasen [1984]), perhaps to signify the author's view that the schizophrenic brain is not, as it were, in "good working order" and needs to be "fixed". More eloquently, Heinrichs (2001) describes schizophrenia as the product of a "neurochemical tempest": "Is schizophrenia really a kind of biological tempest, where tides of neurotransmitters crest and recede? Do substances with cryptic and unpronounceable names play havoc with patches of protein called receptors, and do they upset chemical balances in regions of the brain that control thought, feeling, and movement?" (Ibid., 181). Here, neurotransmission in schizophrenia is likened to a malevolent storm at sea, which reflects, and explains, the uncontrollable and chaotic "storm" of thoughts and feelings associated with schizophrenia. This, again, reflects the view that "all is not as it should be" in the schizophrenic brain.

Such expressions permeate not only literature for popular audiences, but scientific literature on schizophrenia as well. It is unusual to read a scientific article on the biological basis of schizophrenia that does not at some point characterize the schizophrenic brain as beset by a neurobiological "dysfunction", "failure" "disability", "aberrance", "malfunction", "deficit", or "excess". All of these are clearly normative terms: they imply a norm, or standard, in relation to which the activity of the brain is assessed as deviant. Moreover, this deviance is not merely statistical, but normative in the proper sense, because it supports the notion that the schizophrenic brain is malfunctioning or dysfunctional and not just different or unusual. The assumption that the brain of the schizophrenic patient is in some sense not working "as it ought" or "as nature intended", then, is central to biological approaches to psychiatry. Tacitly or explicitly, much biological research in psychiatry is fueled by the desire to identify what, precisely, has "gone wrong" in the schizophrenic brain, or how, precisely, nature has "erred" in those brains.

This assumption appears to be confirmed by the fact that biological research *has* been successful at uncovering diverse biological disparities between the brains of schizophrenic patients and those of normal controls. Although there is no single anomaly that is sufficient or necessary for schizophrenia – that is, there does not exist some biological abnormality that all and only persons with schizophrenia possess (Heinrichs [2001, 249]) – there are nonetheless promising results that suggest that, in at least some cases, some of the abnormal and disparate schizophrenic symptoms such as delusions, hallucinations, disorganized speech or behavior, affective flattening or avolition (APA [2000, 312]), may be associated with genetic and neurobiological abnormalities. Hence, while there may be no unique biological index of schizophrenia, the diverse evidence of biological discrepancies cannot simply be ignored.

However, one cannot validly infer from the claim that a distinct biological abnormality has been discovered among a given subgroup of schizophrenics, to the claim that the abnormality represents a biological *dysfunction*. The fact that something is different or unusual does not mean it is dysfunctional! Left-handedness is statistically unusual but is probably not the result of a dysfunction, even if it has a distinct biological cause. Moreover – as noted in Chapter 4 (Section 4.2.2, under "Malfunctioning and Maladaptive Neural Circuitry") – even if this biological abnormality produces a behavioral or psychological consequence that is maladaptive, unfortunate, or inopportune, this does not mean that it is dysfunctional. It is unfortunate that childbirth is often painful but that does not mean that normal childbirth is caused by a dysfunction of the female reproductive system. Left-handedness is statistically unusual, biologically driven, and negatively valued in many parts of the world, but that does not imply that anything has "gone wrong" in the left-handed individual. By similar reasoning, it is maladaptive for a person to suffer hallucinations or delusions, but that does not mean that the production of hallucinations is the result of some dysfunction. *That* represents a further claim for which independent empirical evidence must be adduced.

More importantly, even if something is unable to perform the function that it has, that does not imply that it is *dysfunctional*. If one binds a person's legs with rope, that person's legs will be unable to perform their natural function of walking – but that does not mean those legs are dysfunctional. Rather, they are simply prevented from performing their function by unusual environmental circumstances. More generally, for any given biological entity, *X*, and any function, *F*, the functional status of *X* with respect to *F* does not merely fall under one of two categories – functional or dysfunctional – but rather, one of *four* different categories: *X* has the function *F* and is capable of performing this function (functioning properly); *X* has the function *F* but is unable to perform *F* due

to an abnormal environment (unable to function due to abnormal environment); *X* has the function *F* but is unable to perform *F* where this inability is not due to an abnormal environment (dysfunctional); and *X* does not have the function *F* (without function). All four of these possibilities will be examined in more detail using a simple neurobiological example (Section 5.1). This chapter will argue that the existence of biological abnormalities associated with schizophrenia does not imply that something in the schizophrenic brain is specifically dysfunctional, rather than functioning properly or unable to function due to an abnormal environment.

This chapter will be divided into two sections. The first section (Section 5.1) will provide an analytic schema that defines the "four categories of functioning" noted in the previous paragraph. It will use the fairly straightforward example of the vertebrate neuromuscular junction (NMJ) to show how given abnormalities and diseases of the NMJ can illustrate all four of these functional categories. This classification will be essential for evaluating the evidence from schizophrenia research.

The second section (Section 5.2) will begin with a brief introduction to schizophrenia, and then will examine two different, fairly popular neurobiological approaches to the etiology of schizophrenia, the neurochemical approach and the neurodevelopmental approach. The neurochemical approach to schizophrenia seeks the root abnormality of schizophrenia in neurotransmitter systems such as the dopamine system or in the abnormal relation between different such systems. Specifically, the dopamine hypothesis of schizophrenia will be examined, according to which certain symptoms of schizophrenia are the result of the abnormally high availability of dopamine. It will be argued that even in those cases in which such an abnormality exists, it is probably wrong to claim that the *dopamine system* itself is dysfunctional with respect to the regulation of dopamine. Rather, it may represent a case in which the dopamine

system is unable to perform its function due to an abnormal environment that is associated, potentially, with an abnormality in a different transmitter system. This implies that the fact that a biological system is abnormal does not mean that it is dysfunctional.[146]

The neurodevelopmental approach to schizophrenia seeks the root cause of schizophrenia in an early developmental abnormality, such as a pregnancy or birth complication, or an abnormality in an ongoing developmental process, that interferes with normal development in such a way as to produce unusual neurobiological characteristics. Various versions of the neurodevelopmental hypothesis of schizophrenia will be examined. It will argue that on some versions, it would be *wrong* to say that schizophrenia stems from a neurodevelopmental dysfunction. Rather, one would have to say that in the case of schizophrenia, developmental abnormalities in the brain represent a plastic response of the brain to an abnormal formative environment, and not a dysfunction. However, on other versions of the hypothesis, it would be *correct* to say that schizophrenia stems from a neurodevelopmental dysfunction. Since the correct neurodevelopmental hypothesis is currently unknown (assuming that one of them is correct), then whether or not schizophrenia stems from a neurodevelopmental dysfunction is also currently unknown. This reinforces the point that one cannot validly infer the presence of a biological dysfunction from a biological abnormality.

5.1 FOUR CATEGORIES OF FUNCTIONING

This section will review the definition of "dysfunction" presented in Chapter 3 and show how it permits four different categories of functioning to be defined. For the sake of convenience they will be termed "functional", "unable to function due to abnormal environment", "dysfunctional", and "without function". A fairly simple

[146] See below (Section 5.1), which recapitulates the difference between an entity's being "dysfunctional" and merely being "unable to perform its function due to an abnormal environment".

example using the neuromuscular junction (NMJ) will be given to illustrate each of these categories. This section will allow one to classify the psychiatric examples that will be presented in Section 5.2 accordingly. For example, according to the dopamine hypothesis of schizophrenia, it should allow one to judge whether the dopamine system in the schizophrenic brain is "functional", "unable to function due to an abnormal environment", "dysfunctional", or "without function" with respect to the regulation of dopamine.

To recapitulate the definition of "dysfunction" presented in Chapter 3 (see Section 3.2.2), to say of an individual entity, *X*, that it is dysfunctional with respect to some activity, *F*, means that:

(i) the function of *X* is *F*;

(ii) *X* is not able to perform *F*; and

(iii) if *X* is not in the normal environment for its functioning, then if *X* were in the normal environment for its functioning, *X* would not be able to perform *F*.

As noted there, this definition rests crucially on the notion of a "normal environment for an entity's functioning". The importance of this notion is that something is not dysfunctioning simply because it is unable to perform its proper function, but because it is unable to perform its proper function in the normal environment for its functioning. As defined there, a "normal environment for an entity's functioning" refers to any one out of the range of environments within which the entity's ancestral progenitors[147] performed the activity that currently constitutes its function, and in which those performances were fitness enhancing. For example, if one of the functions of legs is

[147] The "ancestral progenitors" of a token of a trait within an individual refers to the same trait in that individual's ancestors.

to walk, but someone cannot walk because his or her legs are bound with rope but otherwise normal, it would be appropriate to say that the person's legs are not really "dysfunctional", but rather, that they are unable to function because they are in an abnormal environment. The view that the function of an entity is relative to some notion of a "normal environment" is fairly standard in philosophical discussions of etiological function (e.g., Millikan [1984, 33-4]), even if it is left implicit.

On the basis of this definition, at least four categories of functioning can be defined. More precisely, for any given entity, *X*, and a given activity, *F*, one can define four different relationships between *X* and *F*:

(I) *Functional*: *X* has the function *F* and *X* is able to perform *F*;

(II) *Unable to Function Due to Abnormal Environment*: *X* has the function *F* and *X* is not able to perform *F* because *X* is not in the normal environment for its functioning;

(III) *Dysfunctional*: *X* has the function *F*, *X* is not able to perform *F*, and if *X* is not in the normal environment for its functioning, then if *X* were in that environment it would not be able to perform *F*; and

(IV) *Without function*: *X* does not have the function *F*.

As noted in Chapter 3, this definition of "dysfunction" implies that if *X* is dysfunctional with respect to *F*, then there must be some structural difference between *X* and one of *X*'s ancestral progenitors (where this progenitor did perform its function) and that this discrepancy does, or would, prevent *X* from performing *F* in its normal environment. The example of the NMJ will be used to illustrate each of these for categories of functioning. The NMJ is used because, as described in Chapter 4 (Section

4.2.2, under "Synaptic Selection"), the NMJ is a paradigmatic case for the application of neural selectionist theories of synaptic structure formation, along with the development of ocular dominance columns in the mammalian visual cortex. Hence, the NMJ allows one to present an empirically well-confirmed and detailed example for each of the four functional categories.

As discussed in the previous chapter (Section 4.2.2, under "Synaptic Selection"), a selection process brings about the pattern of neural connectivity in the NMJ. This allows one to assign *functions* to synaptic properties of the NMJ. For example, a selection process accounts for the one-to-one pattern of connectivity between motor neurons and skeletal muscle fibers in the rat NMJ by two weeks after birth (see Figure 4.6). Even if the specific mechanisms of this selection process are not known, one can infer that the multiple axons that innervated each muscle fiber at birth varied with respect to some property, and that this variation contributed to their differential retention on that fiber. This implies that once the one-to-one pattern of connectivity is established, some property of the remaining axon has the *function* of ensuring this synaptic connection with the target.

Since synaptic selection is an activity-dependent process, it relies essentially upon neural activation – that is, the production of action potentials or graded potentials in the pre-synaptic neuron and the release of neurotransmitter onto receptors of the post-synaptic target. Hence, it is plausible that one property of the innervating axon in the rat NMJ that explains its differential retention is its capacity to produce an action potential and release the neurotransmitter acetylcholine (ACh) upon the post-synaptic receptors embedded in the muscle fiber, thereby eliciting an excitatory potential in that muscle fiber. Consequently, according to SPE, one can say one of the *functions* of the remaining

innervating axon is to *activate the muscle fiber through the release of ACh onto the post-synaptic receptors*.

The determination of this function of the innervating axon permits one to construct four different scenarios involving the NMJ that illustrate the four different categories of functioning, or the four different relationships between the biological entity, *X – the pre-synaptic neuron* – and the activity, *F – activation of the muscle fiber through the release of ACh onto the post-synaptic receptors*. It is obvious that, whatever else it may include, the normal environment for the functioning of the pre-synaptic neuron must be an environment in which the following two normal input/output conditions hold:

> (1) the dendrites of that pre-synaptic neuron are themselves innervated by an incoming axon (input); and
> (2) that pre-synaptic neuron synapses onto ACh receptors that are capable of producing excitatory potentials within the muscle fiber (output).

Although there are other conditions that must be imposed on the normal environment for that neuron's functioning (e.g., adequate amounts of extracellular calcium and potassium to generate an action potential), these two conditions are necessary features of that normal environment. In other words, if either condition (1) or (2) is not met then the pre-synaptic neuron is not within the normal environment for its functioning, since, historically, conditions (1) and (2) have been instrumental for every prior performance of its function. The four categories of functioning can now be illustrated through the following four scenarios.

Functional. Through the selection process described above, a motor neuron comes to have the function of activating the muscle fiber through the release of ACh onto the

post-synaptic receptor (see Figure 5.1). With respect to those features of its environment that have a bearing on its ability to release ACh, its current environment is similar. Consequently, it can be said to occupy the normal environment for its functioning – one in which it is adequately innervated by incoming neurons, and in which there are adequate and functional post-synaptic receptors in the muscle fiber that it innervates, as well as adequate amounts of extracellular calcium, potassium, and so on. Upon sufficient excitatory stimulation, it produces an action potential and releases ACh onto the muscle fiber, thus producing an excitatory post-synaptic potential in the muscle fiber and thereby completing the performance of its function. This is a paradigm case of synaptic transmission. Here, the neuron is functional with respect to its function.

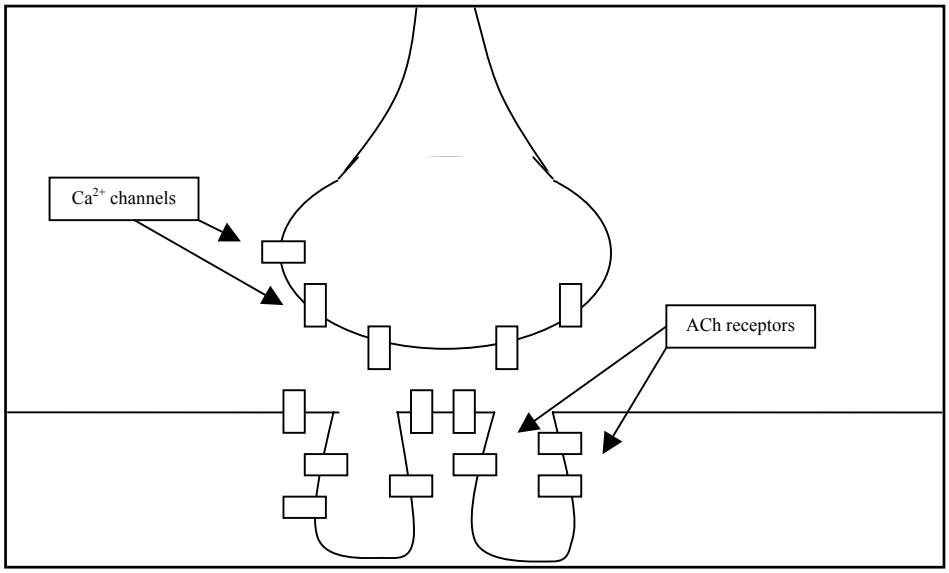

Figure 5.1: Functional motor neuron.

Unable to function due to abnormal environment. The neuron undergoes the same selection process described in the previous paragraph and hence comes to have the function of activating the muscle fiber. However, a muscular disorder, Myasthenia Gravis, affects the ability of the muscle fiber to *receive* transmitter. In Myasthenia Gravis

it is believed that antibodies to the ACh receptor destroy the ability of the muscle fiber to respond to transmitters (Rich *at al*. [1994]). (See Figure 5.2, in which certain postsynaptic receptors are marked with an '**X**' to indicate an inability to receive transmitter.) In this case, the motor neuron is not in the normal environment for its functioning, because one of the essential prerequisites which has allowed it to activate the muscle fiber – namely, the fact that functional ACh receptors are embedded within the muscle fiber – does not exist. (This is represented by the violation of condition [2] above, which implies a disruption of its normal "output" condition.) However, structurally, the motor neuron is similar to one that is able to perform its function. This suggests that were the motor neuron in the normal environment for its functioning – that is, in the absence of the neuromuscular disease that affects the postsynaptic receptors – it would be able to perform its function. Consequently, the motor neuron does not satisfy the third criterion (iii) of the definition of "dysfunction" and cannot be said to be dysfunctional with respect to that activity. Rather, it is *unable to function due to an abnormal environment*. In this case it would be as counterintuitive to say that the motor neuron is "dysfunctional" as it would be to say of a person's legs that are bound with rope that they are "dysfunctional" because they are unable to walk. A comparison of Figure 5.1 and Figure 5.2 illustrates that structurally, the motor neuron is identical in both cases and therefore it should not qualify as "dysfunctional" in one and not the other.

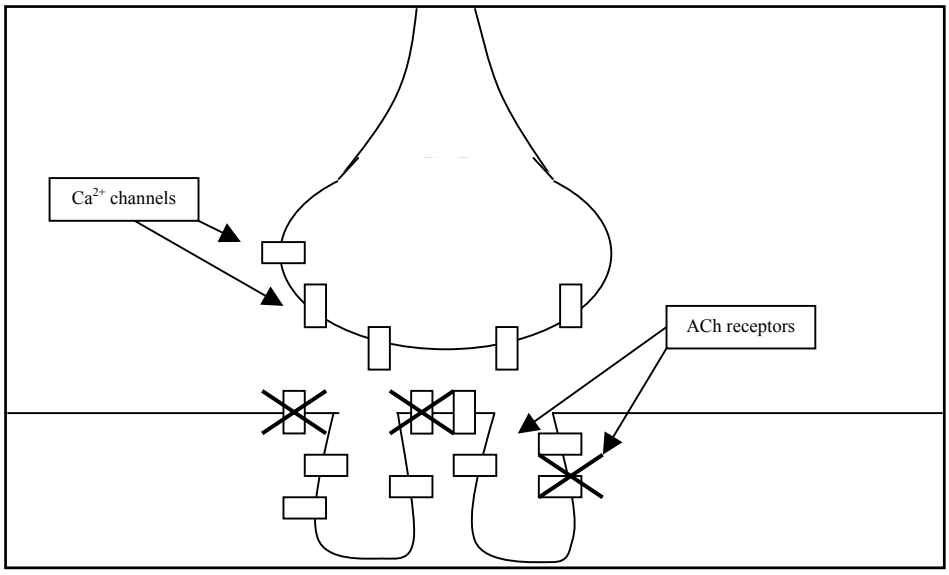

Figure 5.2: Motor neuron that is unable to function due to abnormal environment.

Dysfunctional. The neuron undergoes the same selection process described in the previous two paragraphs and hence comes to have the function of activating the muscle fiber. In most respects its current environment is similar to that in which it came to possess the function, and hence it is in the "normal environment for its functioning". However, due to a neuromuscular disorder such as Lambert-Eaton syndrome the neuron is unable to release neurotransmitters onto the muscle fiber. Lambert-Eaton syndrome is an autoimmune disorder that involves antibodies to the voltage-gated calcium channels in the presynaptic neuron, which are essential for the release of neurotransmitter (Takamori [2004]). Hence, it disrupts a neuron's ability to release ACh. (See Figure 5.3, in which certain calcium channels are marked with an '**X**' to indicate an inability to conduct calcium.) Since it is in its normal environment but cannot perform its function then it is dysfunctional.

One might argue that the neuron afflicted by Lambert-Eaton syndrome is *not* in the normal environment for its functioning, because the presence of calcium channel

antibodies could not have been a component of the environment within which its progenitors were able to perform their function. This can be admitted; the essential point is that, supposing that the disease has degraded the calcium channels in the pre-synaptic terminal in such a way as to prohibit sufficient calcium influx, the neuron's *structure* has been affected to such an extent that it cannot perform its function. Hence, even if one were to place the neuron in the normal environment for its functioning by removing the calcium channel antibodies, it would still be unable to perform its function because of this structural divergence. Consequently, it satisfies all three criteria for the definition of dysfunction.

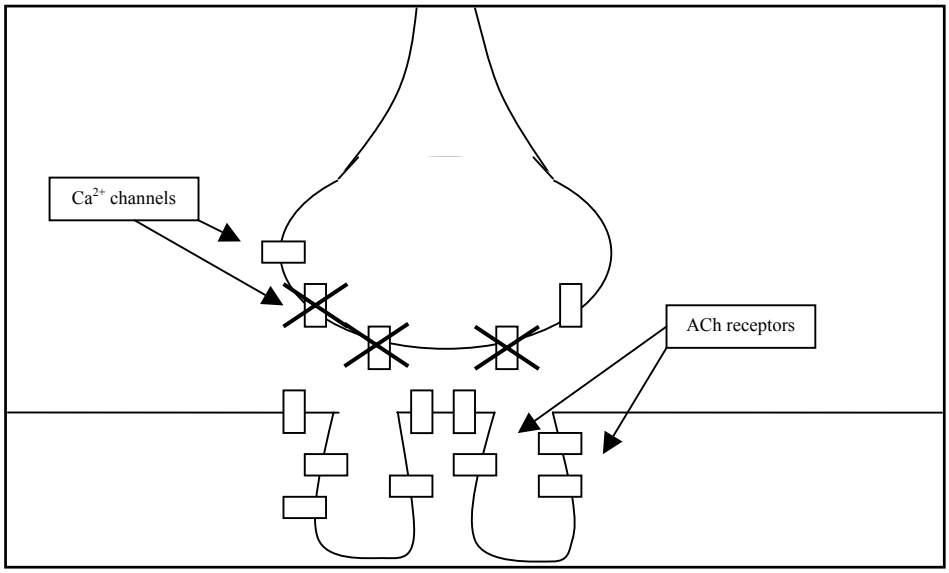

Figure 5.3: Dysfunctional motor neuron.

Without function. As the situation has been described above, the reason that the motor neuron has the function of activating the muscle fiber is because its capacity to activate the muscle fiber is one of the reasons that it was differentially retained on the muscle fiber. As noted in Chapter 4 (Section 4.2.2, under "Evidence for Neural Selection") the selection process that determines the one-to-one pattern of connectivity

between motor neuron and muscle fiber in the rat NMJ is more or less complete by two weeks after birth. Before this selection process begins, then, the multiple motor neurons that innervate the muscle fiber do not have the function of activating the muscle fiber by the release of ACh since they have not yet been selected for because of that capacity.[148] In this case, the motor neuron simply does not have the relevant function at all. This implies that if a neuromuscular disease were to destroy its capacity to release transmitter within the first two weeks of birth, one could not say that the neuron would be dysfunctional.

In many cases, the distinction between "functional" and "unable to function due to an abnormal environment" can only be drawn in a conventional manner, because it relies upon the precise way that one chooses to describe the function of the entity. This is a consequence of a well-known problem of indeterminacy for function ascriptions, which is called the "problem of functional indeterminacy" (Dretske [1986]; Sterelny [1990]; Shapiro [1992]; Goode and Griffiths [1995]; Neander [1995]; Buller [1997]). The heart provides a simple example of this indeterminacy. By beating, the heart circulates blood and thereby sustains the life of the organism; this explains why hearts were selected for by natural selection and hence why hearts currently exist. Consequently, the function of the heart can be described in several ways: one can say that the function of the heart is to *beat*, or that the function of the heart is to *circulate blood*, or that the function of the heart is to *sustain the life of the organism*. No version of the etiological theory of function described in Chapter 4 allows one to "pick out" one of these three descriptions of the heart as the "uniquely correct" description of its function: since they are all equally justified, there is no uniquely correct such description.

[148] Perhaps the motor neurons have been selected for by natural selection operating over an evolutionary time frame because of their capacity to activate muscle fibers, in which case they would have the relevant function even before the synaptic selection process has affected its structure. This will be ignored for the purpose of illustration.

This indeterminacy can create ambiguities in deciding whether the heart is "functioning normally" or "unable to function due to an abnormal environment". For example, suppose that one says that the function of the heart is to *circulate blood*. Suppose then that, due to a severe hemorrhage in the brain, a person's heart is unable to circulate a sufficient quantity of blood to keep that person alive, even though it is still beating and pumping blood. In this case, one would have to say that the heart is unable to perform its function due to an abnormal environment. Alternatively, suppose that one says that the function of the heart is merely to *beat*. In this case, so long as the heart is beating, it is able to perform its function even if, due to a hemorrhage, it is not able to properly circulate the blood. Under this description one would have to say that the heart is functioning normally. Consequently, in the case of this hemorrhage, whether one chooses to say that the heart is "functional", or, alternatively, that it is "unable to perform its function due to an abnormal environment", is a purely conventional decision.

The same can be said of the NMJ described above. There, it was said that the motor neuron has the function of *activating the muscle fiber through the release of ACh onto the post-synaptic receptors*. Consequently, if, due to Mysthenia Gravis, the muscle fiber cannot be activated, one would have to say that the motor neuron is unable to perform its proper function. Alternatively, suppose that one merely attributes to the motor neuron the function of *releasing ACh onto the muscle fiber*. Then, one would have to say that the motor neuron *is* functional in the case of Myasthenia Gravis, since Myasthenia Gravis does not affect the ability of the motor neuron to release neurotransmitter onto muscle fiber (see Figure 5.2). Hence, from the point of view of this dissertation, the distinction between "functional" and "unable to function due to an abnormal environment" is not a very important distinction.

The distinction that *is* crucial to the dissertation is that between the case in which an entity is *dysfunctional* and the case in which it is *not dysfunctional*. The problem of functional indeterminacy does not affect this distinction; hence it is not a matter of convention. In the case of Lambert-Eaton syndrome, which affects the ability of the neuron to produce action potentials (see Figure 5.3), the structure of the neuron is sufficiently affected that it cannot initiate the sequence of events that explains its differential retention on the muscle fiber. In other words, it cannot perform any of its "functions", however described. Similarly, if the heart is unable to pump blood, due to stenosis of the mitral valve, then it cannot perform any of the activities that explain its differential reproduction: pumping, circulating blood, assisting cell metabolism or contributing to the survival of the organism. This is because its structure prohibits it from doing so. As Neander (1995, 120) puts it, it is dysfunctional because it cannot perform the function that is "most specific to the trait in question", or its "most specific function", even within the normal environment for its functioning.

Equipped with this analytical schema, one is now in a position to evaluate the various hypothesized neurochemical and neurodevelopmental mechanisms for the etiology of schizophrenia that will be presented in the next section. This enables one to answer the following question in a precise manner: under what conditions does a given hypothetical mechanism for the etiology of schizophrenia represent a "dysfunctional" mechanism, and under what conditions does it represent a "non-dysfunctional" one? Note that even if it is empirically unknown which hypothesis is *correct*, as long as one is given a sufficiently detailed model for the etiology of schizophrenia, one can classify the mechanisms postulated in that model to be either "dysfunctional" or "non-dysfunctional" – which is all that is necessary for evaluating the question that the dissertation poses.

5.2 NEUROBIOLOGICAL APPROACHES TO SCHIZOPHRENIA

This section will begin by describing some of the symptoms of schizophrenia and characterizing some of the main directions of contemporary research (Section 5.2.1). Then it will present the dopamine hypothesis of schizophrenia, and go on to present evidence suggesting that in the case of schizophrenia, there is little warrant for claiming that the dopamine system itself is dysfunctional, but rather, that it is unable to perform its function due to an abnormal neurochemical environment. This will show that the mere presence of a well-defined biochemical abnormality associated with schizophrenia is not in itself conclusive reason to assume that schizophrenia stems from a biological dysfunction (Section 5.2.2). One might argue, of course, that even if the dopamine system *itself* is not dysfunctional, then the existence of such a gross abnormality entails that there must be *some* dysfunction in the brain that is creating such an abnormal environment for the dopamine system. The final part of the section will respond to this claim by using a neurodevelopmental example to show that the neurobiological abnormality in question could represent a plastic response of the brain to a relatively unusual environment (Section 5.2.3). In this case, nothing "inside" the person would be dysfunctional. Since there is currently not enough evidence to rule this hypothesis out, then there is not enough evidence to rule out the possibility that schizophrenia does not stem from a biological dysfunction at all.

5.2.1 A Brief Overview of Schizophrenia

This section will provide a short introduction to the topic of schizophrenia, which will be useful for the remainder of the section, which evaluates different proposed etiological mechanisms. The reason that the dissertation specifically focuses on schizophrenia is that this illness – or, perhaps, the heterogeneous set of conditions that fall under that category – tends to represent the paradigm of lay or colloquial usage of

"mental disorder" itself, or its more pejorative associations: "madness", "insanity", and "craziness" (e.g., see Smith [1982, 13]; Heinrichs [2001, 3]). Secondly, it cannot be doubted that schizophrenia is a central target of biologically-oriented research in psychiatry. For this reason, one cannot impugn the following argument on the grounds that it exploits a marginal diagnostic category, or one that is on the fringes of biological research. Consequently, this section purports to show that a condition that is currently considered to be a central target of biologically-oriented research in psychiatry, as well as a paradigm case of "mental disorder", does not in any obvious way stem from a biological dysfunction.

Schizophrenia as a Diagnostic Construct

Ever since the inception of "dementia praecox" as a diagnostic category by the German psychiatrist Emil Kraepelin in the fifth edition of his *Psychiatrie* (1896, 426-41), schizophrenia has been characterized by a diversity of symptoms that do not always overlap within the same person. In the sixth edition of 1899, Kraepelin himself suggested the existence of three different subtypes of schizophrenia, the hebephrenic, the catatonic, and the paranoid types. The paranoid type is characterized by the rather "florid" symptoms of hallucinations and delusions – those symptoms most paradigmatically tied to "madness" or "insanity". The catatonic form is marked by stupor, the maintenance of rigid and often bizarre postures, negativism or automatic obedience, and occasionally, outbursts of reckless activity. Finally, Kraepelin describes the hebephrenic form in the seventh edition of his textbook as "the gradual or subacute development of a simple more or less profound mental deterioration" (Kraepelin [1981 (1907), 230]).[149] Although this form may be marked by hallucinations or delusions, it is primarily characterized by

[149] The above quote is from the English translation of the seventh edition of his textbook, which was originally published in 1907.

emotional indifference, as well as a progressive disorganization of thought, speech, and behavior (Ibid., 234). Today this is referred to as the "disorganized" subtype (APA [2000, 314-5]).

Eugen Bleuler, who in 1911 coined the term "schizophrenia" to replace the descriptively inaccurate expression "dementia praecox"[150], used the term to characterize a "group of psychoses" (Bleuler [1950 (1911), 7]), all of which are characterized by "a more or less clear-cut splitting of the psychic functions" (Ibid., 9). Bleuler retained Kraepelin's three subtypes, and added a fourth, "schizophrenia simplex" (Ibid., 235). This subtype was expressed by "mildly pathological symptoms" (Ibid., 239), and a gradual and generalized decline in intellectual and emotional functioning. Thus, heterogeneity at the syndromal level has always been a recognized feature of schizophrenia.

Kraepelin's classification of the major subtypes of schizophrenia has been remarkably preserved through the present century. These subtypes are clearly seen in the contemporary DSM (APA [2000]) and ICD (WHO [1992]) mental disorder classifications.[151] According to the most recent edition of DSM, the DSM-IV-TR (APA [2000]), in order to possess schizophrenia one must usually have at least two of the following five symptoms:

(1) delusions
(2) hallucinations
(3) disorganized speech (e.g., frequent derailment or incoherence)
(4) grossly disorganized or catatonic behavior
(5) negative symptoms, i.e., affective flattening, alogia, or avolition (Ibid., 312)

[150] Schizophrenia, Bleuler noted, is not always associated with early onset nor with an irreversible mental deterioration as implied by Kraepelin's designation (Bleuler [1950 (1911), 7]).
[151] See *fn*. 1 on the DSM classification of mental disorders, and *fn*. 54 on the ICD classification.

The DSM classification also explicitly utilizes the essentially Kraepelinean subtypes of paranoid (Ibid., 313-314), catatonic (Ibid., 315-316), and disorganized (Ibid., 314-315), the latter of which is identified in the text with Kraepelin's hebephrenic subtype (Ibid., 314).[152] Kraepelin's subtypes, then, still play an influential role in diagnosis. Moreover, factor analysis, a statistical technique that can be used for grouping symptoms into clusters, suggests three main syndromes: "psychomotor poverty", "disorganization", and "reality distortion", which correspond, at least in part, to the three main types codified in the DSM (Liddle [1987]). Thus, contrary to the rather pessimistic conclusion drawn by Boyle (1990) – who claims that the diagnostic category has been so mutable over the course of the twentieth century that it is both ontologically suspect and practically useless – schizophrenia as a diagnostic construct has been a fairly stable one.

Another common way of arranging these symptoms is by classifying them under two main categories, "Type I" and "Type II" syndromes (Crow [1980a; 1980b]). The original purpose of this classification was to sharpen an earlier distinction between so-called "positive" and "negative" symptoms of schizophrenia (Strauss *et al*. [1974]) and to correlate these two different types of symptoms with different neurobiological abnormalities. In short, Type I or "positive" symptoms of schizophrenia refer to the presence of florid abnormalities such as hallucinations and delusions; the Type II or "negative" symptoms of schizophrenia refer to the "absence" of expected cognitive or behavioral traits such as affective flatness, social withdrawal, lack of volition, or catatonia. However, this distinction has also been criticized, because many patients exhibit mixed positive and negative symptoms, and the supposed etiological clarification gained by the distinction has not received adequate confirmation (Zuckerman [1999, 325-

[152] In addition, the DSM-IV-TR includes an "undifferentiated" subtype (Ibid., 316) which meets the criteria for schizophrenia but does not meet specific criteria for the first three subtypes, and a "residual" type (Ibid., 316-317), in which symptoms are present in an attenuated manner. The ICD-10 classification is essentially similar (WHO [1992]).

331]). Additionally, some symptoms, such as those associated with disorganized thought, do not fall obviously under either category (Jablensky [2001, 19-20]; Walker and Lewine [1988, 316-7]).

Neurobiological Approaches to Schizophrenia

Schizophrenia appears to affect 0.5% - 1.5% of the general population (APA [2000, 308]), prevalence rates being comparable across the world, with prognosis being somewhat better in less industrialized countries (Sartorius *et al*. [1986]). It is widely believed to be influenced genetically (Moises and Gottesman [2001]; Tsuang [2000]), although it does not follow a classic Mendelian pattern of inheritance, and attempts to localize specific genes have not been successfully replicated (Riley and McGuffin [2000]; Torrey and Yolken [2000]; Moldin [1997]). Neurobiological approaches have fared somewhat better. The "cardinal" neurological abnormalities associated with schizophrenia are decreased cerebral cortex volume, lateral ventricle enlargement, and enlarged sulci (Weinberger [1984]; Woods *et al*. [1996]); these have been found at the onset of the illness, signifying that it is not a secondary effect of the disorder. A recent meta-analysis of neurobiological data accumulated from 1980 to 1999, however, suggests an estimated distribution overlap of about 75% between normal and schizophrenic populations on the specific trait of reduced frontal brain volume (Heinrichs [2001, 109]). What this means is that only about 25% of schizophrenics and non-schizophrenics can be distinguished by utilizing this measure alone.[153] Consequently, although these features

[153] This meta-analysis is over 39 studies, which utilize data derived from both computerized axial tomography and magnetic resonance imaging studies. A meta-analysis published in 1990 (Raz and Raz [1990]) gave a more optimistic estimate, with an estimated distribution overlap of 60%, implying that 40% of schizophrenic patients and normal controls could be distinguished in terms of frontal brain volume. However, the earlier meta-analysis was based exclusively on computerized axial tomography, which is less accurate than magnetic resonance imaging and probably led to an inflated estimation of the actual degree of non-overlap (Heinrichs [2001, 108-110]).

fall far short of the specificity required to constitute neurobiological "markers" for schizophrenia, they may designate a "subtype" of schizophrenia.

The brain regions that are centrally implicated in schizophrenia and that receive substantial research attention are the frontal and temporal lobes (Heinrichs [2001, 106-116, 137-148]). Gross neurostructural abnormalities, subtle neurostructural abnormalities, gross neurofunctional abnormalities, and subtle neurofunctional abnormalities have been claimed for these regions.[154] Yet many of the specific findings tend to be limited in their generality, and have been difficult to replicate consistently (Ibid. [2001, 116-118, 148-149]).[155] An example of a gross neurofunctional abnormality is that described by the "hypofrontality" hypothesis (Ingvar and Franzen [1974]; see Heinrichs [2001, 112-116] for an overview of research) according to which people with schizophrenia have reduced frontal brain activity relative to normal controls. This hypothesis stems from functional imaging studies that purport to show that in some schizophrenic patients there is little difference in prefrontal activity immediately before, and during, the performance of certain cognitive tasks, such as the Wisconsin Card Sort Task, which challenges working memory (Weinberger *et al*. [1986]; Andreasen *et al*. [1992]; Andreasen *et al*. [1997]).[156] However, results have been inconsistent (Heinrichs [2001, 113]; Ingvar [1987]; Andreasen *et al*. [1997]), leading some to suggest abandoning the hypofrontality hypothesis altogether (Gur and Gur [1995]). A meta-analysis over 21 studies suggests that hypofrontality distinguishes about 40% of schizophrenic and non-schizophrenic

---

[154] An example of a gross neurostructural abnormality is ventricle enlargement; an example of a more subtle neurostructural abnormality is neuronal disarray in certain areas of the hippocampus. An example of a gross neurofunctional abnormality is reduced activity in the prefrontal cortex; an example of a more subtle abnormality would be excessive dopamine availability in the mesolimbic dopamine tract.

[155] The following discussion of neurobiological abnormalities is not intended to be a comprehensive overview (see Heinrichs [2001] or Harrison [1999]); rather, it presents certain central findings for the sake of illustrating major trends in schizophrenia research.

[156] Andreasen *et al*. (1992) reports decreased activity only in patients with negative symptoms; however, Gur and Gur (1995) claim that the results have been found over different schizophrenic subtypes.

populations – still a substantial minority – but there exists significant inconsistency between individual studies (Heinrichs [2001, 113]). An example of a more subtle neurofunctional abnormality would be found in abnormal neurotransmission, a hypothesis that will be described below (Section 5.2.2).

An example of a gross neurostructural abnormality would be ventricular enlargement, a hypothesis described earlier. An example of a subtle neurostructural abnormality associated with schizophrenia is provided by postmortem brain research, which has shown microscopic cytoarchitectural disturbances in the prefrontal cortex, hippocampus, entorhinal area, and cingulate gyrus (Beckmann [2001]). According to one hypothesis that stems from this research, schizophrenia is associated with abnormalities in the orientation of pyramidal cells (or "cell disarray") in the hippocampus (Beckmann [2001, 84-5]; Heinrichs [2001, 197-99]). However, there is substantial variation between individual studies, ranging from virtually no overlap between schizophrenic and normal populations to virtually complete overlap (Arnold and Trojanowski [1996, 223]). Meta-analysis over eight studies produces an average overlap of about 49% (Heinrichs [2001, 198]), which suggests that hippocampal abnormalities affect about half of schizophrenic patients. However, the effects of neuroleptic medication may compromise these results (Ibid.). Another neurostructural finding is that of an abnormal distribution of neurons in the white matter directly beneath the prefrontal cortex (Akbarian *et al*. 1993a; 1996]). This anomaly supports a neurodevelopmental theory of schizophrenia since the nature of the distribution suggests a disruption of early cell migration. However, it is based on a relatively small postmortem sample size (Harrison [1999, 602-3]).

Schizophrenia as a Heterogeneous Illness

One oft-cited reason for the often partial and inconsistent research results is that schizophrenia is a "heterogeneous" illness (Crow [1995]; Goldberg and Weinberger

[1995]; Cardno and Farmer [1995]; Tsuang and Faraone [1995]). What this means is that different etiological mechanisms produce the symptom-clusters characteristic of schizophrenia, and therefore that, to the extent that diseases are individuated by pathology, schizophrenia is best conceived as a family of diverse illnesses. The heterogeneity of schizophrenia has given rise to the attempt to correlate the diverse biological abnormalities more precisely with specific "subtypes" of schizophrenia, as noted above. For example, the distinction between Type I and Type II symptoms represented such an attempt (see above, under "Schizophrenia as a Diagnostic Construct"). Its originator suggested that Type I syndrome predicts potential response to neuroleptic drugs – and hence is caused by subtle neurofunctional abnormalities – and that Type II predicts increased ventricle size and poor outcome – and hence is associated with gross neurostructural abnormalities (Crow [1980b, 383-4]). However, as noted above, these results have not held up to empirical scrutiny. Moreover, and unfortunately, the vast majority of schizophrenia research has been conducted on the basis of diagnosis and not on the basis of differentiated subtypes or symptoms (Jablensky [2001, 28]); in other words, research groups are usually composed of people diagnosed with schizophrenia as such, without regard to subtype. Consequently, little systematic work has been done in correlating neurobiological abnormalities with specific subtypes of schizophrenia. The resolution of the problem of heterogeneity, then, will clearly involve some dialectical interplay between biological research results and refined clinical descriptions and classification (Andreasen [1987]).

However, the fact that schizophrenia is probably a heterogeneous illness helps to place the following discussion in proper perspective. The following subsection will examine the dopamine hypothesis as well as the view that the primary neurochemical abnormality in schizophrenia is related to the glutamate system (Section 5.2.2). There is

no implication, however, that abnormalities in glutamatergic transmission underlie most, or even a majority, of diagnoses of schizophrenia. It is enough for the present purposes if abnormal glutamatergic transmission is associated with at least some cases of schizophrenia. This allows one to ask the following question: *For the specific subgroup of persons with schizophrenia for which, e.g., the glutamate hypothesis is valid, or for that specific subtype of schizophrenia which is associated with abnormal glutamatergic transmission, is there warrant for saying that schizophrenia stems from an internal dysfunction*? Hence, the discussion of specific etiological mechanisms will be abstracted away from any specific claim about the prevalence of that mechanism. Whether or not "schizophrenia", as such, turns out to be one illness or many is irrelevant to the discussion.

5.2.2 A Neurochemical Approach: The Dopamine Hypothesis of Schizophrenia

This section will adopt a neurochemical approach to schizophrenia, and, in particular, it will examine a specific neurochemical hypothesis, the dopamine hypothesis. According to the classic formulation of the dopamine hypothesis, schizophrenia – or at least some subtype of schizophrenia (Crow [1980a]) – stems from the overproduction of dopamine in the brain. This section will look at more recent evidence that although such an abnormality may exist, it is more reasonable to suppose that this abnormality stems from an abnormality in the glutamate system, which regulates the dopamine system. Consequently, it would be incorrect to say that in the case of schizophrenia, the dopamine system is dysfunctional; rather, one should say that it is unable to perform its proper function due to an abnormal environment. The next subsection will examine more fully the nature of the glutamate abnormality in question and present the possibility that it results from a plastic brain response to an abnormal formative environment and not an inherent dysfunction. If this is so, then this chapter will have provided at least a plausible

scenario according to which, despite the neurobiological abnormalities associated with schizophrenia, nothing in the brain is dysfunctional. Since, empirically, there is not sufficient warrant for deciding between the various scenarios, there is not sufficient warrant for claiming that schizophrenia stems from a neurobiological dysfunction.

According to one of the initial formulations of the dopamine hypothesis (Carlsson [1974]), schizophrenia stems from excessive production or availability of dopamine in the brain. Two major lines of reasoning contribute to the plausibility of this hypothesis (see, e.g., Pliszka [2003, 232-239] for an overview). The first is that all known antipsychotic ("neuroleptic") medications block the dopamine D_2 receptor (Meltzer and Deutch [1999, 1060])[157]. This is true of the more recent, "atypical" antipsychotics such as clozapine and risperidone – which usually exhibit higher antagonist effects at serotonin (5-HT) receptors relative to their antagonist effects at dopamine D_2 receptors – in addition to the "typical" antipsychotics such as chlorpromazine and haloperidol (Kapur and Seeman [2001]).[158] A second line of evidence is that amphetamines, which can mimic some of the positive symptoms of schizophrenia, act by releasing dopamine (Grace [1991, 2]).

Although few attempts at integrative neurocognitive models for schizophrenia exist – that is, attempts to build models that explain the relation between, e.g., excessive dopamine availability and the characteristic symptoms of schizophrenia[159] – the anatomical distribution of dopamine producing regions, and their terminal fields, are suggestive of basic neurocognitive functions that would be affected.[160]

---

[157] The sole exception to this generalization is reserpine, which depletes vesicular stores of dopamine and is not currently in use (Ibid.).

[158] The atypical antipsychotics also appear to bind the D_3 and D_4 receptors more effectively than D_2 receptors (Kandel *et al*. [2000, 1200]).

[159] See Andreasen (1997) for an overview of various attempts; also see Kapur (2003) for a specific attempt to link the dopamine hypothesis to cognitive disturbance in schizophrenia.

[160] Much of the following discussion of neuroanatomy is based on Pliszka (2003, 66-68).

There are two major groups of dopamine-producing neurons in the brain, both of which are located in the midbrain. The first resides in the ventral tegmental area of the brain and sends groups of axons into two main regions, the limbic system and the cortex, particularly the frontal lobes. The former tract is called the "mesolimbic" dopamine tract because it originates in the midbrain and terminates in the limbic system – those parts of the brain specifically implicated in the regulation of emotion and motivation. One of the most important terminal fields of the mesolimbic tract is the nucleus accumbens of the ventral striatum, which is involved in reward and pleasure (Berridge and Robinson [1998]; Wise [2002]). The latter tract is called the "mesocortical" tract because it originates in the midbrain and terminates in the cortex, especially in the prefrontal cortex. This area of the brain is thought to be implicated in the temporal organization of behavior, motivation, and attention (Kandel *et al*. [2000, 1202]). Consequently, abnormal dopamine production in the ventral tegmental area (VTA) can profoundly affect emotion and cognition.

The second major group of dopamine-producing cells in the brain is in the substantia nigra compacta. The axons from these cells terminate in the neostriatum, which is implicated in (among other things) motor control; it is called the "nigrostriatal" tract because it originates in the substantia nigra and terminates in the striatum. It is believed that insufficient production of dopamine in this region is responsible for the loss of motor regulation found in Parkinson's disease. Consequently, the antagonistic effect of antipsychotic drugs at dopamine receptors in this tract is probably responsible for some of the extrapyramidal side-effects associated with typical antipsychotics, such as the dyskinesias (various forms of motor impairment). The improvement in extrapyramidal side-effects associated with the atypical antipsychotics is probably a result of their

reduced affinity for the dopamine receptor as compared to the typical antipsychotics (Kapur and Seeman [2001, 361]).

However, despite the initial plausibility of the dopamine hypothesis, it has not been verified more directly and faces several challenges (Pliszka [2003, 233]; Grace [2000, 331]). Firstly, if schizophrenia is associated with increased dopamine availability, then there ought to be a measurable increase in homovanillic acid (HVA) in the cerebrospinal fluid (CSF), since HVA is the primary metabolite of dopamine (Pliszka [2003, 233]). But such a correlation has not been consistently discovered (e.g., Post *et al*. [1975]; van Kammen *et al*. [1986]; Beuger *et al*. [1996]). Secondly, although the antagonistic effects of antipsychotics take place more or less immediately, their therapeutic effect can take several weeks (Grace *et al*. [1997, 31]). Consequently, dopamine receptor antagonism alone cannot be sufficient for resolving schizophrenic symptoms.[161] Thirdly, drugs such as lysergic acid diethylamide (LSD) and phencyclidine (PCP) can also produce psychotic symptoms, although neither operates primarily through releasing dopamine: LSD appears to be a serotonin agonist, and PCP, a glutamate antagonist which binds the glutamatergic NMDA receptor. Hence, excess dopamine production does not appear to be necessary for producing schizophrenic symptoms (Pliszka [2003, 233]). Finally, as noted above, the atypical antipsychotics exhibit a decreased antagonist effect on dopamine than the typical antipsychotics, relative to their effect on the serotonin and norepinephrine systems (Ibid., 233-235). This has substantially broadened the scope of recent neuropharmacological research into

[161] This challenge can be accounted for by the hypothesis that the therapeutic action of antipsychotic drugs is not dopamine receptor antagonism as such – which takes place immediately upon application – but a phenomenon called "depolarization block", which refers to the inactivation of dopamine neurons that takes place after several weeks of constant antagonism (Grace *et al*. [1997]).

schizophrenia (e.g., see Holden [2003]), and suggested that other transmitter systems may be deeply implicated.[162]

Moreover, for the present purposes, even if schizophrenia – or at least some subtype of schizophrenia – is related to high dopamine production, it does not follow that the dopamine system is *itself* dysfunctional. One possibility, which will not be considered here, is that in schizophrenia, the dopamine system merely occupies the higher end of its normal functioning. A second possibility, which is more important from the perspective of the dissertation, is that it may be unable to function properly owing to an abnormal environment. This would be the case if, for example, the excessive production of dopamine is a consequence of an abnormality in a different neurotransmitter system. An examination of the structure and function of a typical dopamine-producing neuron can help to clarify the various scenarios that may be at work (see Figure 5.4).[163]

Dopamine is synthesized from tyrosine, which is converted into L-dopa by the enzyme tyrosine hydroxylase (step 1); L-dopa, in turn, yields dopamine through the action of a second enzyme, dopa-decarboxylase (step 2). The newly synthesized dopamine is taken up by a vesicle transport protein into presynaptic vesicles, where it is stored while the neuron is inactive (step 3). Upon activation by an action potential, calcium enters the axon terminal through voltage-gated calcium channels, and this influx of calcium causes the vesicle to release its contents into the synapse (step 4). Here, it can stimulate dopamine receptors on the post-synaptic terminal before it is rapidly removed

---

[162] However, it has recently been argued influentially that the therapeutic effect of atypical antipsychotics such as clozapine is not its multireceptor profile, but rather its relatively fast rate of dissociation from the dopamine D_2 receptor (Kapur and Seeman [2001]; Kapur and Remington [2001]). According to their argument, this dissociation rate attenuates the phasic release of dopamine, but because it does not completely block dopamine activity it leads to a reduced incidence of extrapyramidal effects. This is suggested by the fact that serotonin receptor occupancy alone is not sufficient for atypical antipsychotic effect (Kapur and Seeman [2001, 362]), nor is it necessary; for example, the antipsychotic amisulpride is a selective dopamine receptor antagonist which produces atypical antipsychotic effects but does not block serotonin receptors.

[163] Much of the following material is from Wilcox *et al*. (1999, 3-11) and Kandel *et al*. (2000, 1200-1203).

from the synapse. Some neurotransmitter is degraded by enzymes and some is taken back up into the presynaptic terminal by membrane transporters (step 5). Once inside the terminal it is either degraded by a different enzyme, monoamine oxidase, or recycled and stored in vesicles by vesicle transporters (step 6).

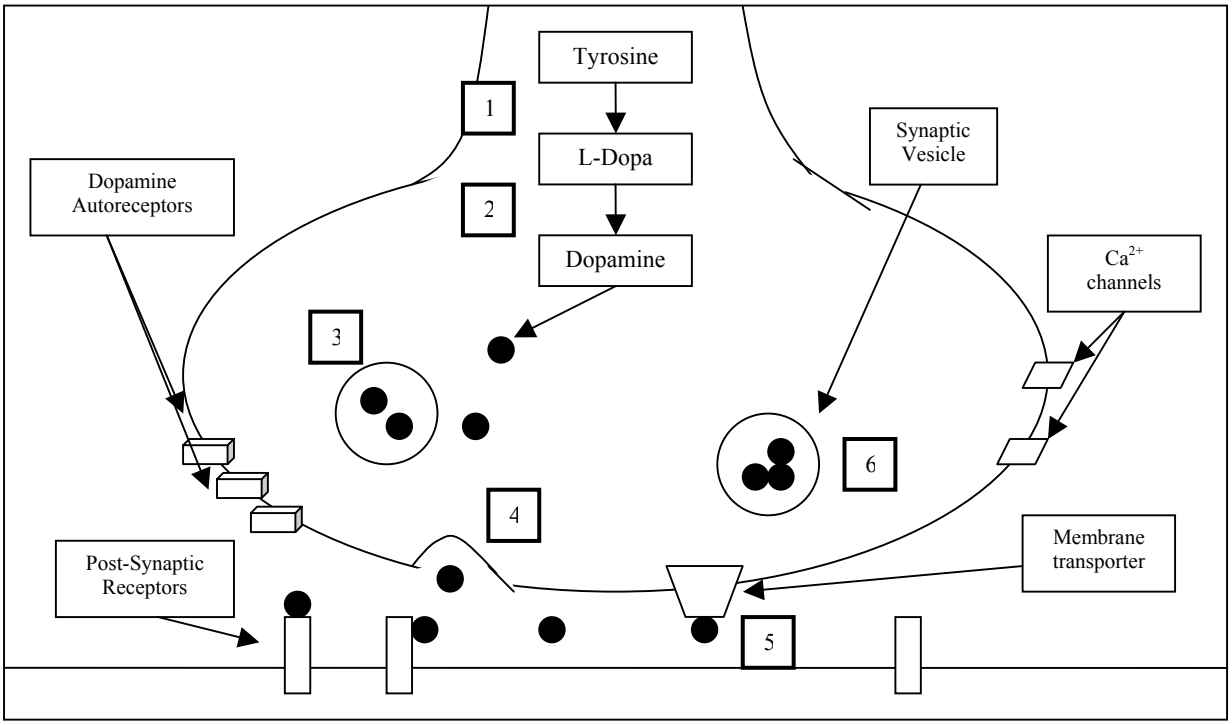

Figure 5.4: Dopamine synapse (redrawn from Kandel *et al*. [2000, 1202]).

The dopamine neuron utilizes a homeostatic mechanism that ensures the presence of a constant level of dopamine in the synapse (Wilcox *et al*. [1999, 6-7]). In other words, when the level of dopamine in the synapse exceeds a certain "set point", the dopamine neuron reduces the amount of dopamine it releases into the synapse; when it drops below this point, it increases the amount of dopamine it releases, thereby maintaining a constant

amount. In the ventral striatum, this level is estimated to lie between 20nM and 50nM (Grace [2000, 336]). It is as if the dopamine neuron is continuously "monitoring" the amount of dopamine in the synapse, and continually adjusting the rate of dopamine synthesis and release in order to maintain this level. In mesolimbic and nigrostriatal dopamine neurons, this synaptic level is monitored by dopamine autoreceptors, that is, receptors that are located on the axon terminal itself (see Figure 5.4).[164] Like the post-synaptic receptors, the dopamine autoreceptors bind synaptic dopamine. When the amount of synaptic dopamine exceeds this "set point", a larger percentage of autoreceptors are bound. In response, the dopamine neuron exercises an inhibitory effect on tyrosine hydroxylase, which is one of the two enzymes crucial for the synthesis of dopamine (step 1). This decreases the available presynaptic pool of dopamine, and hence leads to a reduction in synaptic dopamine back to this "set point". Conversely, when the amount of synaptic dopamine drops below this point, a smaller percentage of autoreceptors are bound, and as a consequence, tyrosine hydroxylase is activated and more dopamine is synthesized. This causes a greater amount of dopamine to be released into the synapse and hence returns the level of synaptic dopamine to this "set point".

On the basis of the structure and function of the typical dopamine neuron, it seems reasonable to infer that *at least one major function* of the dopamine neuron is to dynamically maintain the level of synaptic dopamine at this "set point". Ideally, this statement about the function of the dopamine receptor could be verified by presenting the selection history that explains its current persistence. In other words, it would be ideal if one could present the evolutionary history of the dopamine receptor, or even the ontogenetic history of the dopamine neuron in a single individual, and show that the evolution or persistence of that structure was partly due to a selection process.

[164] Of the five dopamine receptors, D_2 and D_3 function as pre-synaptic autoreceptors as well as post-synaptic heteroceptors (Kandel *et al*. [2000, 1200]).

Unfortunately, it is not known whether or not the dopamine neuron was specifically selected for by natural selection or synaptic selection because it maintained a constant level of synaptic dopamine. However, the apparently homeostatic, self-regulatory structure of the neuron suggests that the neuron has been adapted specifically for this function by a selection process operating over an evolutionary time-scale, an ontogenetic time-scale, or both. At any rate, that a function of the dopamine receptor is to maintain a constant level of synaptic dopamine will be a working assumption for the remainder of this section.

Once a function has been assigned to the dopamine neuron *as a whole*, one can reasonably infer the function of the individual *parts* of the dopamine neuron by assessing the manner in which they contribute to the functioning of the whole. If the function of the whole neuron is to maintain a constant amount of synaptic dopamine, then it is not unreasonable to suggest that the function of each part of that neuron consists in its specific contribution to that global function. For example, knowing that dopamine binding at the autoreceptor has an important regulatory influence on tyrosine hydroxylase, and that this regulatory influence is critical in allowing the dopamine neuron to perform its function of maintaining a constant level of synaptic dopamine, one can reasonably assume that the specific *function* of the dopamine autoreceptor is to regulate the activation of tyrosine hydroxylase. By a similar inference it could be said that the *function* of tyrosine hydroxylase is to regulate the presynaptic availability of dopamine (through the conversion of tyrosine into L-dopa) and that the *function* of the vesicle transporter is to ensure the preservation and storage of presynaptic dopamine by transporting dopamine into the vesicle.[165]

[165] As noted above (Section 5.1.1), because of the problem of functional indeterminacy, there are numerous ways that one might describe the function of any given part of the neuron, and each description is equally correct. For example, one might say that the function of tyrosine hydroxylase is to *convert tyrosine into L-dopa*, or that the function of tyrosine hydroxylase is to *regulate the presynaptic availability of dopamine*, or

Equipped now with a working model of the dopamine neuron, one can understand the diversity of ways in which the regulation of dopamine can be pharmacologically manipulated to produce an excessive amount of synaptic dopamine. This can be used to generate several different hypotheses that would explain the dopamine abnormality. There are several agents that would produce this effect (Kandel *et al*. [2000, 1202]; see Figure 5.4). For example, amphetamines appear to increase the amount of synaptic dopamine by stimulating the process of vesicular release (step 4) (Wilcox *et al*. [1999, 6]), whereas cocaine increases synaptic dopamine by preventing the reuptake of dopamine by the membrane transporter (step 5) (Ibid., 8). A common therapeutic intervention for Parkinson's disease is the administration of the dopamine precursor L-dopa, which increases the presynaptic availability of dopamine (step 2) (Kandel *et al*. [2000, 862]).

Given the spectrum of ways in which the normal activity of the dopamine system can be pharmacologically manipulated, one can envision different specific mechanisms that may underlie the excessive availability of dopamine envisioned by the dopamine hypothesis. For each proposed mechanism, one can then say whether or not the dopamine system would be "dysfunctional", or rather, whether it would be merely unable to function owing to an abnormal environment. Three different scenarios will be described:

> (i) a genetic mutation leads to an amino acid substitution at the dopamine autoreceptor, which, in turn, prevents that receptor from binding dopamine and thereby regulating the further synthesis of dopamine. In this case, the autoreceptor would clearly be unable to perform its function. Moreover, since the mutation

---

that the function of tyrosine hydroxylase is to *contribute to maintaining a constant level of synaptic dopamine*. Fortunately, for present purposes, it is immaterial which description is selected, because, as will be seen immediately below, the conditions under which tyrosine hydroxylase can be said to be *dysfunctional* do not depend on which specific description is selected.

affects the very structure of the autoreceptor in such a way that, even if it is in the normal environment for its functioning, it is prohibited from carrying out that function, then according to the definition presented above it would be *dysfunctional*. If such a scenario were validated empirically, then it would be appropriate to say that at least in some cases, schizophrenia stems from a biological dysfunction on the part of the dopamine neuron, and more specifically, of the receptor;

(ii) due to the administration of cocaine – which, as noted above, inhibits the reuptake of dopamine and thereby increases its synaptic availability – the membrane transporter is prevented from performing its normal function of dopamine uptake. Clearly, the membrane transporter is unable to perform its proper function. However, in this case, the dopamine neuron is not in the normal environment for its functioning. Consequently, in the absence of any significant structural change induced by cocaine exposure, if the dopamine neuron were returned to its normal environment then it would be able to perform its function. In the case of cocaine administration, then, one would have to say that the dopamine neuron is *not* dysfunctional, but rather, that it is unable to perform its normal function due to an abnormal environment;[166]

[166] A slightly different example would yield a case in which the dopamine neuron is dysfunctional. The drug reserpine inhibits the storage of dopamine by interfering with the vesicle transport protein. The presence of reserpine eventually causes long-term structural damage to the vesicles, as a consequence of which, even if the drug is removed, the vesicles are permanently unable to store transmitter (Kandel *et al*. [2000, 1202]). Once the structural damage is accomplished then the dopamine neuron could be called dysfunctional, since even if it were placed in its normal environment (that is, in the absence of reserpine) it would still be unable to store transmitter for structural reasons.

(iii) a third possibility stems from considering not only the way the dopamine neuron is regulated internally, but externally. Dopamine activity in the mesolimbic tract is regulated in part by glutamate neurons in the prefrontal cortex (PFC), which can affect dopamine activity in complex ways (Carlsson and Carlsson [1990, 273-274]; Grace [1991, 6-8]). In this case, abnormal glutamatergic activity could bring about abnormal dopamine activity, as a consequence of which, one could not say that the dopamine system itself is dysfunctional but rather that, due to abnormal glutamate activity, it is unable to perform its proper function due to an abnormal environment.

Several lines of evidence, in fact, suggest this last possibility to be the case. Grace (2000; also see Grace [1991]) reviews evidence that, like dopamine neurons originating in the ventral tegmental area, glutamate neurons originating in the PFC also terminate in the nucleus accumbens and modulate dopamine activity there. Specifically, he provides a model according to which decreased PFC glutamate activity ("glutamate hypofunction") would lead to *increased* dopamine release in the nucleus accumbens ("dopamine hyperfunction") (Grace [2000, 335-6]; also see Grace [1991, 7-8]).[167] In this case, as Grace points out, the primary neurochemical abnormality in schizophirenia would not be in the dopamine system, but in the glutamate system: "[C]urrent models into the pathophysiology of schizophrenia suggest…that the [dopamine] system may be relatively normal, but is subjected to a dysregulation as a consequence of the abnormal control by

[167] Briefly, according to Grace's model, PFC glutamatergic inputs to the nucleus accumbens regulate the tonic level of dopamine, that is, the ability of the dopamine neuron to maintain a constant level of dopamine in the synapse. A decrease in glutamate activity would bring about a decrease in tonic dopamine activity (this would be equivalent to decreasing the "set point" of synaptic dopamine). As a compensatory response to this decreased amount, there is an increase in dopamine synthesis (owing to the decreased autoreceptor binding) and an upregulation of post-synaptic dopamine receptors. As a consequence, normal activity-dependent (phasic) dopamine release would have an abnormally large effect, thus creating a dopamine hyperfunction (Grace [1991, 7]).

cortical glutamatergic afferents…"(Grace [2000, 332]). Carlsson (2001; also see Carlsson and Carlsson [1990]) – who originally proposed the dopamine hypothesis for schizophrenia – arrives at a similar conclusion, arguing that "the elevated dopamine function in schizophrenia could perhaps even be a compensatory, exaggerated, and deleterious response to a failure of a different system" (Carlson [2001, 4]).

Some of the lines of evidence that suggest the presence of glutamate hypofunction in schizophrenia have already been touched upon in this section. First, as noted above, psychostimulants such as PCP inhibit the NMDA glutamate receptor and produce symptoms analogous to schizophrenia; in fact, it has been argued that PCP intoxication bears a closer resemblance to schizophrenia than that due to LSD or amphetamines because it more closely approximates negative as well as positive symptoms (Olney *et al*. [1999, 524]). Second, evidence for "hypofrontality", that is, for decrease prefrontal activity in schizophrenia, would also indicate a relative decrease in glutamate activity in that region, as would evidence for decreased frontal lobe volume in schizophrenia (Carlsson and Carlsson [1990, 275]). Third, some reports suggest that administration of glycine, a transmitter that is necessary for NMDA glutamate receptor function, alleviates schizophrenic symptoms (Meltzer and Deutch [1999, 1067]). Finally, prolonged NMDA receptor antagonism can lead to toxic pyramidal cell excitation in the cortex and limbic system, the areas most typically tied to aberrant cell morphology in schizophrenia (Hirsch *et al*. [1997, 798]; also see Olney *et al*. [1999, 524-5]).

According to the concept of function developed in this dissertation, if the abnormally high levels of synaptic dopamine in the nucleus accumbens represent a response to a glutamatergic abnormality, then one would have to say that the dopamine system is unable to perform its function due to an abnormal environment, rather than that the dopamine system is dysfunctional. This is because, in the absence of substantial

structural alteration in the dopamine system produced by prolonged exposure to an abnormal environment, *if* the mesolimbic dopamine neurons *were* situated within the normal environment for their functioning, then they *would* be able to perform their proper function. This is not to say that there is *no* neurobiological dysfunction in schizophrenia, or that the glutamate hypothesis is ultimately correct. The point is that, since there are multiple plausible scenarios that may underlie the dopamine abnormality, and only some of them implicate dopamine system dysfunction, then one is not warranted in saying that schizophrenia stems from a dysfunction in that system.

One might argue that this conclusion, while valid, is relatively trivial, since it surely implies that even if the *dopamine* system is not dysfunctional in schizophrenia, there must exist *some* neurobiological dysfunction that explains the dopamine abnormality. It may be, for example, that the *glutamate* system is dysfunctional in schizophrenia, or, at least, whatever it is in the brain that explains the abnormal glutamate activity. This possibility will be pursued in the following subsection (Section 5.2.3). However, one plausible scenario that will be presented there is that the glutamate abnormality associated with schizophrenia is the result of a neurodevelopmental abnormality, specifically, an abnormally long or intensive phase of synaptic elimination ("pruning") during childhood and adolescence. Since it is known that the extent and duration of synaptic pruning can represent a plastic response of the brain to differing levels of environmental stimulation, then it is possible that the glutamatergic abnormality in the brain is simply the expression of a plastic response of the brain to a formative environment that is unusually stimulating (or impoverished) in certain respects. If this is so, then it would be inappropriate to say that schizophrenia stems from a biological – or even an "internal" – dysfunction *at all*. Rather, the abnormality is one that is associated with the formative external environment and not, ultimately, with the *internal milieu* of

the person with schizophrenia. Again, the point is not that this scenario is ultimately correct, but that it is plausible, and as a consequence, one is not warranted in inferring the existence of a biological dysfunction merely on the basis of a neurobiological abnormality associated with schizophrenia.

5.2.3 A Neurodevelopmental Approach: Synaptic Selection and Neural Plasticity

The following section examines the neurodevelopmental hypothesis of schizophrenia, and provides a brief overview of different specific models. In particular, it will examine the view that the primary neurodevelopmental abnormality in schizophrenia involves an abnormality in synaptic selection ("pruning") and that this underlies the proposed glutamatergic abnormality described in the previous section. It will argue that, depending on the specific mechanism involved, this pruning abnormality may or may not reflect a developmental dysfunction. Instead of dysfunction, it may, in fact, represent a plastic response of the brain to a formative environment that is unusually stimulating, or impoverished, in certain respects. In this case, the neurobiological abnormalities wrought by this developmental abnormality could not be said to represent a biological, or internal, dysfunction. Since the correct hypothesis is unknown, it is not known whether or not schizophrenia stems from a biological dysfunction. The conclusion arrived at in this subsection, then, reinforces that drawn in the previous subsection, namely, just because schizophrenia may be associated with a biological abnormality does not imply that it stems from a biological dysfunction.

In its broadest form, the neurodevelopmental hypothesis states that schizophrenia results from abnormal genetic or epigenetic events which disrupt early brain development. These early events may interact with later developmental processes to result in the neurobiological abnormalities associated with schizophrenia (see McClure and Weinberger [2001] for a summary statement and overview of the

neurodevelopmental hypothesis). The neurodevelopmental approach probably represents the most prominent current viewpoint on the etiology of schizophrenia (Harrison [1999, 606]).

There are two primary strands of evidence which support the neurodevelopmental hypothesis (Harrison [1997]). The first is that the gross neurostructural abnormalities associated with (at least some subtypes) of schizophrenia, such as ventricular enlargement or reduced frontal lobe volume, appear to be present at the onset of the illness, and are rarely progressive in character. Secondly, some of the more subtle cytoarchitectural abnormalities, such as hippocampal cellular disarray, or abnormal distribution of neurons in the PFC, suggest an early neurodevelopmental origin. This is supported by the absence of observable gliosis – a reactive product of neurodegeneration – in the schizophrenic brain. In other words, if the cellular abnormalities associated with schizophrenia were due to late-onset neurodegeneration then one would expect to find gliosis. Therefore, the absence of gliosis indirectly supports the neurodevelopmental model.

Specific neurodevelopmental hypotheses fall under two types of neurodevelopmental models, an *early* model and a *late* model (Keshavan *et al*. [1994, 240]; Lewis [1997, 386]). The first model emphasizes early neurodevelopmental abnormalities in the etiology of schizophrenia, such as a prenatal lesion in the developing brain, which then interacts with normal neurobiological development to yield the schizophrenic pathology (Murray and Lewis [1987]; Weinberger [1987]). The second model emphasizes deviation in ongoing maturational processes in cortical development rather than any specific early developmental insult.

Most proponents of the early model believe that this early developmental insult occurs during the second trimester of pregnancy. This is because profound

neurostructural abnormalities would result if this insult afflicts the fetus in the first trimester, and gliosis would be apparent if it afflicts the fetus during or after the third trimester (Harrison [1999, 606]). Additional evidence stems from animal models. Induced hippocampal lesions in neonatal rats lead to certain behavioral alterations in the mature rat. These alterations include an increased sensitivity to stress and amphetamines, and hyperlocomotion. However, the induction of similar lesions in mature rats does not produce those alterations (discussed in Grace [2000, 332]). This shows that an early lesion in the developing brain could be responsible for fairly late-onset behavioral abnormalities. Since schizophrenia most frequently emerges in the third decade of life, any neurodevelopmental model must be able to account for the relative quiescence of symptoms during childhood and early adolescence (Harrison [1997, 285-6]).

It has been argued in the literature that the source of this proposed early lesion in schizophrenia may be fetal malnutrition, obstetric complications, or maternal exposure to a virus (McClure and Weinberger [2001, 29-30]). It is also commonly argued that this early insult may be due to a failure of neural migration or an abnormality in programmed cell death (Akbarian *et al*. [1993a; 1993b; 1996]; Jones [1995]; Bunney *et al*. [1997]). Since neural cell migration is a process which ends by the middle of the second trimester, it falls squarely within the postulated "window of vulnerability". A central strand of evidence for the particular hypothesis of abnormal cell migration stems from observations of an abnormal distribution of a certain type of neuron, that expressing the nicotinamide-adenine dinucleotide phosphate-diaphorase (NADPH-d) enzyme, in the white matter directly below the prefrontal and temporal cortices. Relative to normal controls, in schizophrenic subjects there is a decreased density of NADPH-d neurons immediately below the cortex, and an increased density in deeper layers of white matter. This suggests that in the process of early neural migration, these neurons "failed" to reach

their proper destination and were prematurely set in deeper white matter (Akbarian *et al*. [1993a, 175]; Akbarian *et al*. [1993b, 183-4]).[168] It has also been suggested that this abnormal distribution is due to abnormalities in programmed cell death (Akbarian *et al*. [1996, 433]; Bunney *et al*. [1997, 168]).

However, Harrison (1999, 606) criticizes the early model, arguing that schizophrenia has been associated with diverse types of cytoarchitectural abnormalities, and only some of them implicate early (e.g., second trimester) abnormalities. Moreover, the abnormalities that are typically invoked to support the early model – such as those that implicate cell migration – are precisely those that have not been firmly established and replicated by ongoing research. Those cytoarchitectural abnormalities that are more well-established could be due to ongoing developmental processes that are not limited to such a short, specific window of time, such as an abnormality in the normal process of cell adhesion, myelination, or synaptic elimination (synaptic pruning) (Ibid., 606-7). These cytoarchitectural abnormalities include reduced neuronal size, and synaptic loss, in various regions including the PFC (Ibid.; also see McGlashan and Hoffman [2000, 638]; Lewis [1997, 385-6]). Hence, though the evidence still implicates a neurodevelopmental abnormality, it is not necessarily one that accords with an early model. The remainder of the section will examine evidence for a specific version of the late model, namely, the hypothesis that schizophrenia stems from ongoing abnormalities in the process of synaptic pruning, since it is a topic that has already been introduced and reviewed in the previous chapter (Section 4.2).

There are three main lines of evidence that suggest that schizophrenia stems from an abnormality in synaptic pruning. The first is that it would explain the average age of

[168] See Akbarian *et al*. (1996), however, which relies on a larger population than Akbarian *et al*. (1993a; 1993b), and where these results are less prevalent. Also see Harrison (1999, 602-603) for criticism of these results.

onset. Feinberg (1982/83) – who first suggested the pruning hypothesis – reasoned on the basis of evidence that synaptic pruning in the frontal cortex continues until adolescence (Huttenlocher [1979]; also see Rakic *et al*. [1986], which confirms this finding). Since the onset of schizophrenia frequently occurs in adolescence or later, Feinberg (1982/83, 331) suggested that it could result from an abnormality in synaptic pruning, although he remained agnostic about the exact nature of this abnormality: "As a result of some abnormality in this process, too many, too few or the wrong synapses are eliminated (Regrettably, we have no basis to choose among these possibilities)" (Ibid.).

A second line of evidence stems from the fact that synaptic pruning in the PFC appears to affect disproportionately asymmetric type synapses, which are primarily glutamatergic inputs to pyramidal cells (Keshavan [1994, 241]; also see Lewis [1997, 390]). This suggests that abnormally intensive pruning in the PFC could lead to reduced connectivity in glutamatergic neurons or the pyramidal cells they innervate (Keshavan [1994, 252-3]), and that this could manifest itself as glutamate hypofunction, thus providing a basis for the glutamate theory of schizophrenia described in the previous subsection. Postmortem tissue studies have substantiated this inference by finding decreased spine density on PFC pyramidal neurons in schizophrenic brains (See McGlashan and Hoffman [2000, 638] and several references therein). However, Deakin and Simpson (1997), on the basis of postmortem tissue studies, report *increased* levels of glutamate synapses in the prefrontal cortex and speculate that the glutamatergic abnormality in schizophrenia stems from an *arrest* of the normal process of synaptic pruning, rather than an abnormally intensive pruning (Ibid., 288-9).

Thirdly, there exists intriguing neurocomputational evidence that certain symptoms of schizophrenia, in particular, auditory hallucinations, can be generated by excessive synaptic pruning (McGlashan and Hoffman [2000]; also see Hoffman and

McGlashan [1993]). A neural network was constructed that utilizes an initial phase of supervised learning to build a verbal working memory. This allows it to classify degraded inputs as identifiable words. The performance of the network was correlated with the extent of synaptic "pruning", which was simulated by the elimination of synaptic "connections" between neurons. The number of correctly identified words increased as pruning continued until the elimination of about 30% of the network connections had taken place. After this point, performance decreased with continued pruning. After 40% of the connections had been eliminated, the network began generating words during periods of input silence, and producing various forms of speech impairment. The generation of unprovoked words was interpreted as a simulated auditory hallucination (McGlashan and Hoffman [2000, 638]). The authors of the study argue that this model can be used to explain specific symptoms of schizophrenia, age of onset, sex differences in onset and course, specific neurodevelopmental deficits, and degree of cognitive deterioration after onset.

Moreover, as they point out, synaptic pruning may be related in one of several different ways to the pathological reduction of neural connectivity. It may be that the degree of synaptic pruning is relatively normal, but an early developmental insult leads to a reduced initial synaptic density, thus resulting in an abnormally low final synaptic density. Alternatively, it may be that initial synaptic density is relatively normal, but some other factor provokes an unusually intensive or lengthy "window" of synaptic pruning, thus resulting in the same outcome (Ibid., 639). For example, Etienne and Baudry (1990, 42-3) hypothesize that NMDA receptor maturation determines the period of time during which synaptic pruning takes place; thus, any genetic abnormality that produces a delay in NMDA receptor maturation would extend the "window" of pruning and hence result in "overpruning" and consequent PFC abnormalities. Hence, the

neurocomputational model is intended be compatible with both early and late neurodevelopmental models of schizophrenia.

Having provided a specific neurodevelopmental hypothesis for schizophrenia (involvement of abnormal synaptic pruning), one is now in a position to evaluate various specific mechanisms that would give rise to this abnormality and decide, for each proposed mechanism, whether or not that mechanism should be called "dysfunctional" or "non-dysfunctional" – as was done in the previous subsection. Again, these three proposals are merely presented as plausible alternatives; there is no implication that any of them are more warranted than alternate proposals that are not listed here:

> (i) an early developmental lesion afflicts the hippocampal area and thereby causes a severe reduction in synaptic density. As a consequence, although there is no abnormality in the process of synaptic pruning *as such*, the early insult interacts with normal pruning in such as way that a net loss of PFC connectivity results. This, in turn, gives rise to some of the characteristic symptoms of schizophrenia. This possibility is attested to by research indicated earlier (discussed in Grace [2000, 332]), in which hippocampal lesions were induced in postnatal rats, and these lesions gave rise to cytoarchitectural abnormalities in frontal and temporal cortices in adult rats that were associated with behavioral abnormalities. According to Grace (2000, 337), one of the functions of the hippocampus is to inhibit prefrontal inputs to the nucleus accumbens. He speculates that, as a result of this hippocampal lesion, the hippocampus is unable to perform this function. Moreover, since the lesion initially affects only the hippocampus – the environment of which is otherwise normal – one would have to say that the hippocampus is dysfunctional, rather than unable to perform its function due to an

abnormal environment. Consequently, evidence for an early-onset neurodevelopmental model for schizophrenia that implicates a prenatal hippocampal lesion would qualify as evidence that schizophrenia does, in fact, stem from a neurobiological dysfunction;

(ii) there is no early developmental lesion or insult that afflicts fetal brain. Rather, the PFC abnormalities in schizophrenia stem from an abnormally extended "window" of time during which synaptic pruning takes place, thus resulting in an abnormal decrease in synaptic density. Suppose that this window of time is genetically regulated, and that the genetic mechanisms that regulate the length of this window have been selected for by natural selection because they optimize the precise degree of synaptic pruning (McGlashan and Hoffman [2000, 643]).[169] If a genetic mutation causes an abnormal extension in the time period during which pruning takes place, then overpruning may result, thereby disrupting the important cognitive functions that optimal pruning allows. In this case, one would say that a genetically-induced dysfunction is responsible for schizophrenia. This is because the genetic mutation prevents that gene segment from performing its function of regulating the degree of pruning, even when that gene segment is in the normal environment for its functioning. For example, as noted above, Etienne and Baudry (1990, 42-3) hypothesize that NMDA receptor maturation determines the length of the process of synaptic pruning; thus, any genetic abnormality that delays the onset of NMDA receptor maturation would extend the "window" of pruning and

[169] As McGlashan and Hoffman (2000, 643) suggest, synaptic pruning can increase the accuracy, efficiency, and degree of learning; thus, there may have been selection for the maximal degree of synaptic pruning that is compatible with these traits. Consequently, if schizophrenia is due to overpruning, then one could say that natural selection itself inherently tends to produce a risk for schizophrenia. It is possible, of course, that this could lead to dysfunction.

result in “overpruning”. According to this hypothesis one would be able to say that schizophrenia stems from a biological, and specifically genetic, dysfunction on the part of the individual;

(iii) there is no early developmental lesion or insult that afflicts fetal brain. Rather, the glutamate abnormality stems from an abnormally long or intensive window of synaptic pruning. However, rather than being the result of a genetic dysfunction (as suggested in scenario [ii] above) the intensive degree of synaptic pruning represents a plastic response of the brain to a formative environment which is in certain respects unusually overstimulating or impoverished. This last option is suggested by the fact that the nature and degree of synaptic pruning is related to the nature and degree of environmental stimulation. This option will be explored in more detail below. What is crucial is that, according to this scenario, one could not necessarily say that schizophrenia stems from a biological dysfunction at all, but rather, that certain parts of the brain may be unable to perform their proper function due to an abnormal environment, *or* that they are, in fact, functioning normally.

As noted in Chapter 4 (Section 4.2), the nature and extent of synaptic elimination in the cortex is related to the nature and degree of certain types of environmental stimulation. For example, visual activity is a determinant of the extent to which synaptic selection processes “sculpt” the mammalian visual cortex, since complete dark rearing of kittens inhibits that activity altogether. Moreover, the fact that synaptic selection allows the brain to respond in a plastic manner to environmental stimulation is shown by the fact that monocular occlusion (e.g., suture of a single eyelid) causes the vast majority of

neurons in the visual cortex to be selectively responsive only to the non-occluded eye. This is a plastic response of the visual system to abnormally impoverished visual experience. The relation between synaptic elimination and neural activity has also been shown for the neuromuscular junction: the inhibition of motor neuron activity by tetrodotoxin, a sodium-channel blocker, decreases the rate at which synapses are eliminated, while more enhanced activity increases this rate (Purves and Lichtman [1980, 156]). Hence there is a systematic dependence between the nature and degree of synaptic pruning and the nature and extent of neural stimulation.

The existence of this systematic dependence is also suggested by reports that the extent of pruning in the forebrain of domestic chicks is related to emotional experience (Bock and Braun [1998]). In newborn chicks, the extent of dendritic spine reduction in the neostriatum was correlated with the degree of experience with an imprinting situation (which consisted of a mother surrogate and an imprinting tone, followed by behavioral tests). Interestingly, there was little difference in dendritic spine density in chicks exposed only to the auditory tone – in the absence of the emotional content associated with a mother surrogate – and chicks that were not exposed to any stimulus at all. Experience with the mother surrogate, then, both initiated and shaped the course of synaptic pruning in chicks. Is it overly speculative to suggest on this basis that abnormal emotional experience in humans could result in an abnormal degree of synaptic pruning (Ibid., 25)?

Moreover, as noted in Chapter 4, the extent and nature of synaptic elimination is relevant to the manner in which *functions* are attributed to synaptic structures. For example, under conditions of complete dark rearing, ocular dominance columns do not form. However, one cannot say in this case that the visual cortical neurons, or their ocular connections, are "dysfunctional". This is because the very process that allows one to

assign functions to the specific synaptic structures that underlie ocular dominance columns – namely, synaptic selection – has not taken place (see Section 5.1 for a similar discussion concerning the neuromuscular junction). Consequently, one would have to say that the visual cortex of the dark-reared kitten is simply unable to perform its function of mediating binocular vision due to an abnormal environment, or that it lacks that function altogether.

On similar grounds, under conditions of monocular occlusion, almost all of the visual cortical neurons become exclusively responsive to stimulation from the non-occluded eye. This is certainly abnormal, but it does not in any sense represent a "dysfunction" on the part of the visual cortex. Rather, there are two alternate possibilities. First, as in the case of dark-rearing, one might say that the visual cortex of the kitten subjected to monocular occlusion is unable to perform its function of mediating binocular vision due to an abnormal environment. Second, one could say that, if the function of the visual cortex is to maximize visual discrimination *given* the particular environment within which the kitten is raised, the abnormal formation of the visual cortex under conditions of monocular occlusion represents the normal or proper functioning of the visual cortex, rather than an inability to function due to an abnormal environment. As noted above (Section 5.1), in certain cases, whether an entity is unable to perform its function due to an abnormal environment, or whether it is, instead, functioning normally, depends on the precise way in which the function of that entity is described. Since there are often multiple correct descriptions of an entity's function (due to the problem of functional indeterminacy) then the precise way in which the function of an entity is described is often a matter of convention. Consequently, in some cases, whether the entity should be said to be functioning normally, or unable to function owing to an

abnormal environment, is also a matter of convention. Such a case is illustrated by the structure of the visual cortex under conditions of monocular occlusion.

Certain authors who have speculated about a relation between schizophrenia and synaptic pruning have acknowledged the fact that the nature and extent of synaptic pruning is shaped by the nature of environmental stimulation, and have even suggested that the environmental dependence of synaptic pruning may provide a way to model the interaction of "biological" and "environmental" factors in the etiology of schizophrenia. As Keshavan *et al*. (1994) note, one strength of the pruning model of schizophrenia is that the environmental-dependence of synaptic sculpting through pruning would "allow for the integration of psychosocial factors into this pathphysiological model" (Ibid., 257). Moreover, they argue, "In view of the possibility that experience may influence selective survival of certain synapses, it is conceivable that genetic abnormalities of programmed synaptic pruning processes and adverse life experiences in early life could interact to result in pathological brain maturation and consequently the schizophrenic diathesis" (Ibid.). Feinberg (1982/83) himself notes that:

> A late reduction in synaptic processes must not, of course, be blind with respect to their utility: it would hardly do to eliminate heavily used connections and leave unneeded ones intact. Whatever process is involved must be sensitive to the "life experience" of the neurons, i.e., their history of activity. This consideration permits some role for environmental and experiential (including "emotional") factors in the model proposed here for the neuropathology of schizophrenia, as my colleague Simon Auster (personal communication) pointed out. (Ibid., 329)

Finally, as McGlashan and Hoffman (2000) point out, the adaptive significance of neural pruning is related to the fact that "it serves learning by increasing cognitive capacity, accuracy, efficiency, and speed of learning" (Ibid., 643); consequently, the nature and extent of synaptic pruning must be systematically related to the nature and

extent of *what there is to be learned*: in a relatively impoverished environment (e.g., dark rearing) one might expect a relative paucity of synaptic elimination; in a relatively stimulating environment, one might expect a greater degree of synaptic sculpting through some mixture of constructive growth and selective elimination of new synapses.

In summary, what is the relevance of the environmental dependence of the rate, degree, and nature of synaptic pruning to models of schizophrenia that postulate the existence of abnormalities in synaptic pruning in schizophrenia? The general result suggested by several studies is that *differences in the rate and degree of synaptic pruning in different individuals represent the plastic response of different brains to changing and unpredictable environmental circumstances*. Therefore, any specific cytoarchitectural abnormalities resulting from abnormalities in synaptic pruning – such as reductions in glutamatergic connectivity and the potential consequences of glutamate dysfunction for the dopamine system – may represent the brain's "best attempt" to adapt to the degree and nature of those environmental circumstances that participate in sculpting its mature form. This possibility is not proposed here as an ascertained fact; rather, it is suggested as a plausible hypothesis about the nature and function of the process of synaptic elimination itself. However, to the extent that this hypothesis is *plausible*, then one cannot unproblematically infer from the presence of neurobiological abnormalities associated with schizophrenia that schizophrenia stems from a biological dysfunction. The fact that a central target of biological research in psychiatry, as well as a paradigm case of mental disorder, does not necessarily stem from a biological dysfunction, suggests that the view that psychiatric disorders, *in general*, stem from biological dysfunctions, should be treated with suspicion. And this is what this dissertation set out to show.

Chapter 6: Conclusion: A Misbegotten Attempt

If there is a single prescription that emerges from this dissertation, it is a note of caution in the context of psychiatric practice: it advises caution in making the all-too-slippery transition from psychiatric disorders to biological dysfunctions. If careful attention to neurobiological details shows that, in the case of schizophrenia, the biological evidence for the existence of such a dysfunction is ambivalent, then it is reasonable to suppose that careful attention to neurobiological details associated with *other* mental disorders might reveal the same thing. Of course, the dissertation may ultimately be wrong. But this can be considered a strength of the dissertation, rather than a weakness, because it means that the question – "Do psychiatric disorders stem from biological dysfunctions?" – has been translated into a clear, empirical question, rather than an empty slogan.

This conclusion, however – that there is little warrant for the claim that psychiatric disorders stem from biological dysfunctions – should not seem altogether surprising. This is because, as discussed in Chapter 2, the appeal to "internal dysfunctions" was largely motivated by the felt need among biologically-oriented psychiatrists in the early 1970s to justify the medical orientation of their discipline in the face of mounting criticism of the arbitrariness of psychiatric nosology (see Section 2.1.1, under "Homosexuality and the Legitimation Crisis"). Stated simply, an important purpose of the expression was to make psychiatry seem more like physiological medicine, the scientific credentials of which were not generally held in dispute.

However, with few notable exceptions (see Section 2.3), there was little concern with what, precisely, would constitute an internal "dysfunction" and how such "dysfunctions" would be identified. R. E. Kendall, a psychiatrist who eventually

abandoned the attempt to formulate an appropriate definition of "disease" for psychiatry, concluded rightly that the DSM's usage of "dysfunction" does not resolve any fundamental problems, but rather, is "vaguely worded [enough] to allow any term with medical connotations to be either included or excluded in conformity with contemporary medical opinion" (Kendell [1986, 41]). Should it be surprising, then, that given a definition of "dysfunction" that is theoretically defensible and empirically appropriate to the context of psychiatry, such dysfunctions do not clearly emerge from biological research? Or that, in the network of diverse biological anomalies associated with various mental disorders, one cannot necessarily pinpoint any specific dysfunctions?

One might argue that the thesis of the dissertation is relatively trivial. Who cares if mental disorders do not stem from "biological dysfunctions", according to some philosophical definition of "function"? After all, the dissertation has acknowledged evidence of important and substantial biological differences between the brains of people who do, and do not, have schizophrenia. Furthermore, the dissertation does not question that schizophrenia, whatever its basis, is a very horrifying condition – both for the person so afflicted as well as that person's friends and family – and that, hopefully, it will someday be eradicated by the advent of biological and other forms of treatment. It is still a great victory for biological psychiatry that schizophrenia, and other severe mental disorders, can safely be said to stem from biological "abnormalities", or to represent "unfortunate biological conditions", and this dissertation does not seek to deny that victory. So – one might argue – why should it matter whether or not schizophrenia can be said to stem from a "biological dysfunction"?

The reason the thesis matters is that the language of "dysfunction" in psychiatry is powerful and significant. Whether or not the brain can be said to be "dysfunctioning" in the case of a severe mental disorder has a tremendous bearing on the way that mental

disorders are conceptualized in psychiatric practice, as well as among the public. On the surface of it, to say that someone has a mental disorder is often to say that something has gone "wrong", as it were, *inside* the person – inside the person's mind or brain. The idea that nature has in some sense "erred" in the brain of a person with a mental disorder bears, for many, an unmistakable intuitive appeal. It is only natural, then, to want to "look inside" the person – whether through psychodynamic psychology or biological intervention – and find out what "went wrong". The language of inner "dysfunctions", then, supports individualistic models of psychiatry that look "inside" the person rather than "outside" the person to his or her environment.

Suppose, however, that one does not say that mental disorders stem from biological dysfunctions, but rather, one merely states that they have biological "causes", or that they stem from "unusual biological conditions", or simply that the brains of people with mental disorders are "different" from the brains of those without? While these statements may be true, they does not carry the same normative weight *because they are not accompanied by the implicit suggestion that anything in the brain has "gone wrong"*. To say that the neurobiology underlying some forms of schizophrenia reveals an "adaptive response of the brain to an unusual formative environment" simply does not suggest that anything in the brain has "gone wrong", but, in fact, that everything in the brain is going "exactly as it should be" under those circumstances. Taking this perspective to an extreme, one might suggest that the unusual biological formations associated with schizophrenia represent a creative triumph of the human brain to adapt to unusual events. The analogy is that of monocular occlusion in kittens – as noted above (Chapter 5.2.2), the unusual development of the mammalian visual cortex under conditions of monocular occlusion represents a triumph of neurobiological plasticity. It is left to the reader to ponder how conceiving of the biology of mental illness along these

lines might affect his or her conception of these illnesses and of the people who have them.

A second notable prescription emerges from this dissertation, which is targeted specifically at philosophers: it suggests that various attempts to *define* "mental disorder" in terms of a supposed internal dysfunction should be abandoned. Whether or not schizophrenia qualifies as a "mental disorder", according to some well-crafted definition, should not ultimately turn on whether it stems from a biological dysfunction or from a nondysfunctional, plastic response to an unusual environment. Yet, as shown in the last chapter, existing neurobiological evidence is consistent with both of these possibilities.

Yet, does psychiatric research or practice *require* a definition of "mental disorder"? It is sometimes argued that a definition of "mental disorder" is crucial for resolving scientific disputes over controversial diagnostic categories (e.g., Wakefield and First [2003]). For example, should racism be considered a disorder?[170] What about Pre-Menstrual Dysphoric Disorder (APA [2000, 771-774])?[171] Or Self-defeating Personality Disorder (APA [1987, 371-374])?[172] At first glance, it seems obvious that only a clear, agreed-upon definition of "mental disorder" can settle pressing dilemmas such as these. Yet even in the classic example of psychiatric controversy – that concerning the diagnostic status of homosexuality – it is questionable that Spitzer's definition of "mental disorder" actually performed any substantial cognitive service. (See Section 2.1.2, under "Charge to Define Mental Disorder".) This is suggested by the fact that as soon as psychiatrists discovered that the proposed definition actually *excluded* conditions that most psychiatrists thought to be disorders – such as the paraphilias ("sexual deviations") – the definition was promptly discarded. As Spitzer himself observed, many critics

[170] See *fn*. 7.
[171] See Robinson (1998).
[172] See Kutchins and Kirk (1997).

argued that the problem with any proposed "definition" of mental disorder is that it would merely be tinkered with in order to justify, *post hoc*, controversial diagnostic decisions (Spitzer [1978, 16]).

In retrospect, then, it appears rash to have assumed that the problem of demarcation in psychiatric classification would be resolved by invoking any simple biological principle, such as the presence or absence of an internal "dysfunction". Perhaps some fairly straightforward principle *does* exist that will offer a more or less clear, unambiguous, and objective demarcation between psychological conditions that should, and should not, qualify as mental disorders. One such view is that mental disorders involve a psychologically defined "failure of action" (e.g., Fulford [1989]; Bolton and Hill [1996]), rather than a biologically defined "failure of function". According to Bolton and Hill (Ibid., 280), all mental disorders involve what they call a "breakdown of intentionality", although one that does not necessarily involve a disruption of physical functioning. This breakdown can occur when a representational system persistently fails to represent reality correctly, or when it embodies a persistent and unresolved conflict between rules involved in the regulation of action. Many mental disorders probably exhibit both. Yet this definition of disorder, like the biological one, still remains enmeshed in the idea that mental illness is essentially bound up with "failure", whether this failure is located on a biological or psychological level. The project of formulating a robust conceptualization of mental disorder that falls completely outside of the rubric of "failure" remains to be accomplished.

The fact that this dissertation does not advance or endorse any novel conceptualization of mental disorder, or its furtive "essence", means that the conclusion of the dissertation is a largely "negative" or "critical" one, rather than a "constructive"

one. However, it is often essential to backtrack away from a false hypothesis in order to make progress on a correct one.

References

Adams, F. 1979. A Goal-State Theory of Function Attributions. Canadian Journal of Philosophy 9: 493-518.

Agich, G. J. 1994. Evaluative Judgement and Personality Disorder. In Philosophical Perspectives on Psychiatric Diagnostic Classification, edited by J. Z. Sadler, O. P. Wiggins, and M. A. Schwartz, 233-245. Baltimore: Johns Hopkins University Press.

_____. 2002. Implications of a Pragmatic Theory of Disease for the DSMs. In Descriptions and Prescriptions: Values, Mental Disorders, and the DSMs, edited by J. Z. Sadler, 96-113. Baltimore: Johns Hopkins University Press.

Akbarian, S., W. E. Bunney, S. G. Potkin, S. B. Wigal, J. O. Hagman, C. A. Sandman, and E. G. Jones. 1993a. Altered Distribution of Nicotinamide-Adenine Dinucleotide Phosphate-Diaphorase Cells in Frontal Lobe of Schizophrenics Implies Disturbances of Cortical Development. Archives of General Psychiatry 50: 169-177.

Akbarian, S., A. Viñuela, J. J. Kim, S. G. Potkin, W. E. Bunney, and E. G. Jones. 1993b. Distorted Distribution of Nicotinamide-Adenine Dinucleotide Phosphate-Diaphorase Neurons in Termporal Lobe Implies Anomalous Cortical Development. Archives of General Psychiatry 50: 178-187.

Akbarian, S., J. J. Kim, S. G. Potkin, W. P. Hetrick, W. E. Bunney, and E. G. Jones. 1996. Maldistribution of Interstital Neurons in Prefrontal White Matter of the Brains of Schizophrenic Patients. Archives of General Psychiatry 53: 425-436.

Albright, T. D., T. M. Jessell, E. R. Kandel, and M. I. Posner. 2000. Neural Science: A Century of Progress and the Mysteries that Remain. Neuron 25: S1-S55.

Allen, C. and M. Bekoff. 1995a. Biological Function, Adaptation, and Natural Design. Philosophy of Science 62: 609-22.

_____. 1995b. Function, Natural Design, and Animal Behavior: Philosophical and Ethological Considerations. In Perspectives in Ethology, Volume 11: Behavioral Design, edited by N. S. Thompson, 1-46. New York: Plenum Press.

American Psychiatric Association. 1952. Diagnostic and Statistical Manual of Mental Disorders. Washington D.C.: American Psychiatric Association.

_____. 1968. Diagnostic and Statistical Manual of Mental Disorders. 2nd ed. Washington D.C.: American Psychiatric Association.

_____. 1980. Diagnostic and Statistical Manual of Mental Disorders: DSM-III. 3rd ed. Washington D.C.: American Psychiatric Association.

_____. 1987. Diagnostic and Statistical Manual of Mental Disorders: DSM-III-R. 3rd ed. rev. Washington D.C.: American Psychiatric Association.

_____. 1994. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV. 4th ed. Washington D.C.: American Psychiatric Association.

_____. 2000. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR. 4th ed. text rev. Washington D.C.: American Psychiatric Association.

Amundson, R. 1989. The Trials and Tribulations of Selectionist Explanations. In Issues in Evolutionary Epistemology, edited by K. Hahlweg and C. A. Hooker, 413-32. Albany: State University of New York Press.

Amundson, R. and G. V. Lauder. 1994. Function without Purpose: The Uses of Causal Role Function in Evolutionary Biology. Biology and Philosophy 9: 443-69.

Andreasen, N. C. 1984. The Broken Brain: The Biological Revolution in Psychiatry. New York: Harper and Row.

_____. 1987. The Diagnosis of Schizophrenia. Schizophrenia Bulletin 13: 9-22.

_____. 1997. Linking Mind and Brain in the Study of Mental Illnesses: A Project for a Scientific Psychopathology. Science 275: 1586-93.

Andreasen, N. C., P. Nopoulos, D. S. O'Leary, D. D. Miller, T. Wassink, and M. Flaum. 1999. Defining the Phenotype of Schizophrenia: Cognitive Dysmetria and its Neural Mechanisms. Biological Psychiatry 46: 908-20.

Andreasen, N. C., D. S. O'Leary, M. Flaum, P. Nopoulos, G. L. Watkins, L. L. Boles Ponto, and R. D. Hichwa. 1997. Hypofrontality in Schizophrenia: Distributed Dysfunctional Circuits in Neuroleptic-Naïve Patients. Lancet 1730-1734.

Andreasen, N. C., K. Rezai, V. W. Swayze, M. Flaum, P. Kirchner, G. Cohen, and D. S. O'Leary. 1992. Hypofrontality in Neuroleptic Naïve Patients and Patients with Chronic Schizophrenia. Assessment with Xenon Single Photon Emission Computed Tomography and the Tower of London. Archives of General Psychiatry 49: 943-958.

Antonini, A., and M. P. Stryker. 1993a. Rapid Remodeling of Axonal Arbors in the Visual Cortex. Science 260: 1819-1821.

_____. 1993b. Development of Individual Geniculocortical Arbors in Cat Striate Cortex and Effects of Binocular Impulse Blockade. The Journal of Neuroscience 13: 3549-3573.

Arnold, S. E., and J. Q. Trojanowski. 1996. Recent Advances in Defining the Neuropathology of Schizophrenia. Acta Neuropathologica 92: 217-231.

Ausubel, D. P. 1961. Personality Disorder is Disease. American Psychologist 16: 69-74.

Ayala, F. J. 1968. Biology as an Autonomous Science. American Scientist 56: 207-21.

_____. 1970. Teleological Explanations in Evolutionary Biology. Philosophy of Science 37: 1-15.

Ayer, A. J. 1952. Language, Truth, and Logic. New York: Dover.

Bateson, G., D. D. Jackson, J. Haley, and J. Weakland. 1956. Toward a Theory of Schizophrenia. Behavioral Science 1: 251-64.

Bandura, A., D. Ross, and S. A. Ross. 1961. Transmission of Aggression through Imitation of Aggressive Models. Journal of Abnormal and Social Psychology 63: 575-82.

Barlow, H. B. 1988. Neuroscience: A New Era? Nature 331: 571.

Bayer, R. 1981. Homosexuality and American Psychiatry: The Politics of Diagnosis. New York: Basic Books.

Bayer, R., and R. L. Spitzer. 1982. Edited Correspondence on the Status of Homosexuality in DSM-III. Journal of the History of the Behaviuoral Sciences 18: 32-52.

_____. 1985. Neurosis, Psychodynamics, and DSM-III. Archives of General Psychiatry 42: 187-96.

Bechtel, W. 1986. Teleological Function Analyses and the Hierarchical Organization of Nature. In Current Issues in Teleology, edited by N. Rescher, 26-48. Lanham, MD.: University Press of America.

Beckmann, H. 2001. Neuropathology of the Endogenous Psychoses. In Contemporary Psychiatry (Vol. 3), edited by F. Henn, N. Sartorius, H. Helmchen, and H. Lauter, 81-100. Berlin: Springer-Verlag.

Beckner, M. 1959. The Biological Way of Thought. New York: Columbia U.P.

_____. 1969. Function and Teleology. Journal of the History of Biology 2: 151-164.

Bedau, M. 1990. Against Mentalism in Teleology. American Philosophical Quarterly 27: 61-70.

_____. 1991. Can Biological Teleology be Naturalized? Journal of Philosophy 88: 647-55.

_____. 1992. Where's the Good in Teleology? Philosophy and Phenomenological Research 52: 781-805.

_____. 1993. Naturalism and Teleology. In Naturalism: A Critical Appraisal, edited by S. J. Wagner and R. Warner, 23-51. Notre Dame, Indiana: University of Notre Dame.

Berridge, K. C., and T. E. Robinson. 1998. What is the Role of Dopamine in Reward: Hedonic Impact, Reward Learning, or Incentive Salience? Brain Research Reviews 28: 309-369.

Berti, R., Durand, J. P., Becchi, S., Brizzi, R., Keller, N and Ruffat, G. Eye Degeneration in the Blind Cave-Dwelling Fish *Phreatichthys andruzzii*. Canadian Journal of Zoology 79: 1278-85.

Beuger, M., D. P. van Kammen, M. E. Kelley, and J. Yao. 1996. Dopamine Turnover in Schizophrenia Before and After Haloperidol Withdrawal: CSF, Plasma, and Urine Studies. Neuropsychopharmacology 15: 75-86.

Bigelow, J., and R. Pargetter. 1987. Functions. Journal of Philosophy 84: 181-96.

Binnie, C. D. 2003. Cognitive Impairment During Epileptiform Discharges: Is it Ever Justifiable to Treat the EEG? Lancet Neurology 2: 725-30.

Bisti, S., C. Gargini, and L. M. Chalupa. 1998. Journal of Neuroscience 18: 5019-5025.

Black, J. E., and W. T. Greenough. 1986. Induction of Pattern in Neural Structure by Experience: Implications for Cognitive Development. In Advances in Developmental Psychology, Vol. 4, edited by M. E. Lamb, A. L. Brown, and B. Rogoff, 1-50. Hillsdale, N. J.: Lawrence Erlbaum.

_____. 1997. How to Build a Brain: Multiple Memory Systems Have Evolved and Only Some of Them are Constructivist. Behavioral and Brain Sciences 1997: 558-559.

Blashfield, R. K. 1984. The Classification of Psychopathology: Neo-Kraepelinian and Quantitative Approaches. New York: Plenum Press.

Bleuler, E. 1950 (1911). Dementia Praecox; or, the Group of Schizophrenias. New York: International Universities Press.

Bock, J., and K. Braun. 1998. Differential Emotional Experience Leads to Pruning of Dendritic Spines in the Forebrain of Domestic Chicks. Neural Plasticity 6: 17-27.

Bock, W., and G. von Wahlert, G. 1965. Adaptation and the Form-Function Complex. Evolution 19: 269-299.

Bolhuis, J. J. 1994. Neurobiological Analyses of Behavioral Mechanisms in Development. In Causal Mechanisms of Behavioral Development, edited by J. A. Hogan, and J. J. Bolhuis, 16-46. Cambridge: Cambridge University Press.

Bolton, D. 2000. Continuing Commentary: Alternatives to Disorder. Philosophy, Psychiatry, and Philosophy 7:141-153.

_____. 2001. Problems in the Defintion of 'Mental Disorder'. The Philosophical Quarterly 51: 182-199.

_____. 2003. Meaning and Causal Explanations in the Behavioural Sciences. In Nature and Narrative, edited by B. Fulford, K. Morris, J. Sadler, and G. Stanghellini, 113-125. Oxford: Oxford University Press.

Bolton, D., and J. Hill. 1996. Mind, Meaning, and Mental Disorder: The Nature of Causal Explanation in Psychology and Psychiatry. Oxford: Oxford University Press.

Boorse, C. 1975. On the Distinction Between Disease and Illness. Philosophy and Public Affairs 5: 49-68.

_____. 1976. Wright on Functions. Philosophical Review 85: 70-86.

_____. 1977. Health as a theoretical concept. Philosophy of Science 44: 542-73.

_____. 1982. What a Theory of Mental Health Should Be. In Psychiatry and Ethics: Insanity, Rational Autonomy, and Mental Health Care, edited by R. B. Edwards, 29-48. Buffalo, NY: Prometheus Books.

_____. 2002. A Rebuttal on Functions. In Functions: New Essays in the Philosophy of Psychology and Biology, edited by A. Ariew, R. Cummins, and M. Perlman, 63-112. Oxford: Oxford University Press.

Bothwell, M. 1995. Functional Interactions of Neurotrophins and Neurotrophin Receptors. Annual Review of Neuroscience 18: 223-253.

Boyle, M. 1990. Schizophrenia: A Scientific Delusion? London: Routledge.

Braithwaite, R. B. 1953. Scientific Explanation. Cambridge: Cambridge University Press.

Brandon, R. N. 1990. Adaptation and Environment. Princeton, N.J.: Princeton University Press.

Bridgman, P. W. 1936. The Nature of Physical Theory. New York: Dover.

Brown, M. C., J. K. S. Jansen, and D. Van Essen. 1976. Polyneural Innervation of Skeletal Muscle in New-Born Rats and its Elimination During Maturation. Journal of Physiology 261: 387-422.

Buller, D. J. 1997. Individualism and Evolutionary Psychology (or: In Defense of 'Narrow' Functions). Philosophy of Science 64: 74-95.

_____. 1998. Etiological Theories of Function: A Geographical Survey. Biology and Philosophy 13: 505-27.

_____. 2002. Function and Design Revisited. In Functions: New Essays in the Philosophy of Psychology and Biology, edited by A. Ariew, R. Cummins, and M. Perlman, 222-243. Oxford: Oxford University Press.

Bunney, B. G., S. G. Potkin, and W. E. Bunney. 1997. Neuropathological Studies of Brain Tissue in Schizophrenia. Journal of Psychiatric Research 31: 159-173.

Burge, T. 1979. Individualism and the Mental. In Midwest Studies in Philosophy IV, edited by P. A. French, T. E. Uehling, and H. K. Wettstein, 73-121. Minneapolis: University of Minnesota Press.

Buss, D. 1999. Evolutionary Psychology: The New Science of the Mind. Boston: Allyn and Bacon.

Campbell, D. T. 1956. Perception as Substitute Trial and Error. Psychological Review 63: 330-342.

_____. 1960. Blind Variation and Selective Survival as a General Strategy in Knowledge-Processes. In Self-Organizing Systems; Proceedings, edited by M. C. Yovits, and S. Cameron, 205-231. New York: Pergamon Press.

_____. 1974. Evolutionary Epistemology. In The Philosophy of Karl Popper, edited by P. A. Schlipp, 413-463. LaSalle, IL.: Open Court.

_____. 1988. A General 'Selection Theory', as Implemented in Biological Evolution and in Social Belief-Transmission-with-Modification in Science. Biology and Philosophy 3: 171-177.

Canfield, J. 1964. Teleological Explanations in Biology. British Journal for the Philosophy of Science 14: 285-95.

Canguilhem, G. 1991. The Normal and the Pathological. New York: Zone Books.

Cardno, A. G., and A. E. Farmer. 1995. The Case For or Against Heterogeneity in the Etiology of Schizophrenia: The Genetic Evidence. Schizophrenia Research 17: 153-159.

Carlsson, A. 1974. Antipsychotic Drugs and Catecholamine Synapses. Journal of Psychiatric Research 11: 57-64.

_____. 2001. Neurotransmitters – Dopamine and Beyond. In Current Issues in the Pharmacology of Schizophrenia, edited by A. Breier, P. V. Tran, J. M. Herrea, G. D. Tollefson, and F. P. Bymaster, 3-11. Philadelphia: Lippincott Williams & Wilkins.

Carlsson, M., and A Carlsson. 1990. Interactions between Glutamatergic and Monoaminergic Systems within the Basal Ganglia – Implications for Schizophrenia and Parkinson's Disease. Trends in Neurosciences 13: 272-276.

Carnap, R. 1950. Logical Foundations of Probability. Chicago: University of Chicago Press.

Changeux, J. P. 1985. Neuronal Man. New York: Pantheon Books.

_____. 1997. Variation and Selection in Neural Function. Trends in Neurosciences 20: 291-292.

Changeux, J.-P., and A. Danchin. 1976. Selective Stabilization of Developing Synapses as a Mechanism for the Specification of Neuronal Networks. Nature 264: 705-11.

Chevaleyre, V., F. C. Moos, and M. G. Desarmerien. 2002. Interplay between Presynaptic and Postsynaptic Activities is Required for Dendritic Plasticity and Synaptogenesis in the Supraoptic Nucleus. Journal of Neuroscience 22: 265-273.

Clark, D. M. 1986. A Cognitive Approach to Panic Disorder. Behaviour Research and Therapy 24: 461-470.

_____. 1997. Panic Disorder and Social Phobia. In Science and Practice of Cognitive Behaviour Therapy, edited by D. M. Clark, and C. G. Fairburn, 119-153. Oxford: Oxford University Press.

Clarke, P. G. H., and W. M. Cowan. 1975. Ectopic Neurons and Aberrant Development During Neural Development. Proceedings of the National Academy of Sciences of the United States of America 72: 4455-4458.

_____. 1976. The Development of the Isthmo-optic Tract in the Chick, with Special Reference to the Occurrence and Correction of Developmental Errors in the

Location and Ocnnections of Isthmo-optic Neurons. Journal of Comparative Neurology 167: 143-64.

Cohen, S., and R. Levi-Montalcini. 1956. A Nerve Growth Stimulating Factor, Isolated from Snake Venom. Proceedings of the National Academy of Sciences of the United States of America 42: 571-574.

Cooper, D. 1967. Psychiatry and Anti-Psychiatry. London: Tavistock.

Cooper, J. E., R. E. Kendell, B. J. Gurland, et al. 1972. Psychiatric Diagnosis in New York and London. London: Oxford University Press.

Cosmides, L, and J. Tooby. 1999. Toward an Evolutionary Taxonomy of Treatable Conditions. Journal of Abnormal Psychology 108: 453-464.

Cowan, W. M. 1973. Neuronal Death as a Regulative Mechanism in the Control of Cell Number in the Nervous System. In Development and Aging in the Nervous System, edited by M. Rockstein, 19-41. New York: Academic Press.

_____. 1978. Aspects of Neural Development. In Neurophysiology III, edited by R. Porter, 149-191. Baltimore: University Park Press.

Craver, C. F. 2001. Role Functions, Mechanisms, and Hierarchy. Philosophy of Science 68: 53-74.

Crick, F. 1989. Neural Edelmanism. Trends in Neurosciences 12: 240- 248.

Crow, J. F. 1979. Genes the Violate Mendel's Rules. Scientific American 240 (2): 134-46.

Crow, T. J. 1980a. Molecular Pathology of Schizophrenia: More than One Disease Process? British Medical Journal 280: 66-68.

_____. 1980b. Positive and Negative Schizophrenic Symptoms and the Role of Dopamine. II. British Journal of Psychiatry 137: 383-386.

_____. 1995. A Continuum of Psychosis, One Human Gene, and Not Much Else – The Case for Homogeneity. Schizophrenia Research 17: 135-145.

Cummins, K. W. 1988. The Study of Stream Ecosystems: A Functional View. In Concepts of Ecosystem Ecology: A Comparative View, edited by L. R. Pomeroy, and J. J. Alberts, 247-262. New York: Springer-Verlag.

Cummins, R. 1975. Functional Analysis. Journal of Philosophy 72: 741-765.

_____. 1983. The Nature of Psychological Explanation. Cambridge, Mass.: MIT Press.

_____. 2002. Neo-Teleology. In Functions: New Essays in the Philosophy of Psychology and Biology, edited by A. Ariew, R. Cummins, and M. Perlman, 157-172. Oxford: Oxford University Press.

Cziko, G. 1995. Without Miracles: Universal Selection Theory and the Second Darwinian Revolution. Cambridge, Mass.: MIT Press.

Dain, N. 1994. Psychiatry and Anti-Psychiatry in the United States. In Discovering the History of Psychiatry, edited by M. S. Micale and R. Porter, 415-444. Oxford: Oxford University Press.

Darden, L., and J. A. Cain. 1989. Selection Type Theories. Philosophy of Science 56: 106-29.

Darwin, C. 1998 (1859). The Origin of Species. New York: The Modern Library.

Davies, A. M., C. Bandtlow, R. Heumann, S. Korsching, R. Hermann, and H. Thoenen. 1987. Timing and Site of Nerve Growth Factor Synthesis in Developing Skin in Relation to Innervation and Expression of the Receptor. Nature 326: 353-358.

Davies, P. S. 2000. Malfunctions. Biology and Philosophy 15: 19-38.

_____. 2001. Norms of Nature: Naturalism and the Nature of Functions. Cambridge, MA: MIT Press.

Dawkins, R. 1989 (1976). The Selfish Gene. Oxford: Oxford University Press.

Deakin, J. W. F., and M. D. C. Simpson. 1997. A Two-Process Theory of Schizophrenia: Evidence from Studies in Post-Mortem Brain. Journal of Psychiatric Research 31: 277-295.

Detwiler, S. R. 1936. Neuroembryology. An Experimental Study. New York: Macmillan.

Dover, G. 2000. Dear Mr Darwin: Letters on the Evolution of Life and Human Nature. Berkeley: University of California Press.

Dretske, F. 1986. Misrepresentation. In Belief: Form, Content, and Function, edited by R. Bogdan, 17-36. Oxford: Clarendon Press.

_____. 1988. Explaining Behavior: Reasons in a World of Causes. Cambridge, Mass.: MIT Press.

Ebendal, T., L. Olson, A. Seiger, and K. –O. Hedlund. 1980. Nerve Growth Factor in the Rat Iris. Nature 286: 25-8.

Edelman, G. M. 1978. Group Selection and Phasic Reentrant Signaling: A Theory of Higher Brain Function. In The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function, edited by G. M. Edelman and V. B. Mountcastle, 51-100. Cambridge, Mass.: MIT Press.

_____. 1987. Neural Darwinism: The Theory of Neuronal Group Selection. New York: Basic Books.

Edelman, G. M., and L. H. Finkel. 1984. Neuronal Group Selection in the Cerebral Cortex. In Dynamic Aspects of Neocortical Function, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan, 653-695.

Edelman, G. M., and G. Tonini. 2001. Consciousness: How Matter Becomes Imagination. London: Penguin.

Edwards, R. B., ed. 1982. Psychiatry and Ethics: Insanity, Rational Autonomy, and Mental Health Care. Buffalo, NY: Prometheus Books.

Elliott, T., and N. R. Shadbolt. 1997. Neurotrophic Factors, Neural Selectionism, and Neuronal Proliferation. Behavioral and Brain Sciences 20: 561-562.

Ellis, A. 1967. Should Some People be Labeled Mentally Ill? Journal of Consulting Psychology 31: 435-46.

Enc, B., and F. Adams. 1992. Functions and Goal-Directedness. Philosophy of Science 59: 635-654.

Engelhardt, H. T., Jr. 1976. Human Well-Being and Medicine: Some Basic Vale-Judgements in the Biomedical Sciences. In Science Ethics and Medicine, edited by H. T. Engelhardt, Jr., and D. Callahan, 120-139. Hastings-on-Hudson, NY: Hastings Center.

Erde, E. L. 1979. Philosophical Considerations Regarding Defining “Health”, “Disease”, Etc. and their Bearing on Medical Practice. Ethics in Society and Medicine 6: 31-48.

Etienne, P., and M. Baudry. 1990. Role of Excitatory Amino Acid Neurotransmission in Synaptic Plasticity and Pathology. An Integrative Hypothesis Concerning the Pathogenesis and Evolutionary Advantages of Schizophrenia-Related Genes. Journal of Neural Transmission 29 (suppl.): 39-48.

Eysenck, H. J., J. A. Wakefield, and A. F. Friedman. 1983. Diagnosis and Clinical Assessment: The DSM-III. Annual Review of Psychology 34: 167-93.

Faber, R. J. 1984. Feedback, Selection, and Function: A Reductionistic Account of Goal-orientation. In Methodology, Metaphysics and the History of Science, edited by R. S. Cohen and M. W. Wartofsky, 43-135. Dordrecht: D. Reidel.

Fadiman, A. 1997. The Spirit Catches You and You Fall Down. New York: Farrar, Straus and Giroux.

Feighner, J. P., E. Robins, S. B. Guze, et al. 1972. Diagnostic Criteria for Use in Psychiatric Research. Archives of General Psychiatry 26: 57-63.

Feinberg, I. 1982/83. Schizophrenia: Caused By a Fault in Programmed Synaptic Elimination During Adolescence? Journal of Psychiatric Research 17: 319-334.

Fogden, M., and P. Fogden. 1974. Animals and their Colors. New York: Crown Publishers.

Foucault, M. 1967. Madness and Civilization: A History of Insanity in the Age of Reason. New York: New American Library.

Frankfurt, H. G., and Poole, B. 1966. Functional Analyses in Biology. British Journal for the Philosophy of Science 17: 69-72.

Fulford, K. W. M. 1989. Moral Theory and Medical Practice. Cambridge: Cambridge University Press.

_____. 1994. Closet Logics: Hidden Conceptual Elements in the DSM and ICD Classifications of Mental Disorders. In Philosophical Perspectives on Psychiatric Diagnostic Classification, edited by J. Z. Sadler, O. P. Wiggins, and M. A. Schwartz, 211-32. Baltimore: Johns Hopkins University Press.

_____. 1999. Nine Variations and a Coda on the Theme of an Evolutionary Definition of Dysfunction. Journal of Abnormal Psychology 108: 412-20.

_____. 2000. Teleology Without Tears. Philosophy, Psychiatry, & Psychology 7: 77-94.

Garmezy, N. 1978. Never Mind the Psychologists: Is it Good for the Children? The Clinical Psychologist 31: 1-6.

Garson, J. 2005. Function and Teleology. In A Companion to the Philosophy of Biology, edited by S. Sarkar and A. Plutynski, forthcoming. Malden, MA.: Blackwell.

Gaze, R. M. 1974. Neuronal Specificity. British Medical Bulletin 30: 116-121.

Gaze, R. M., M. Jacobson, and G. Szekely. 1963. The Retinotectal Projection in Xenopus with Compound Eyes. Journal of Physiology 165: 484-499.

_____. 1965. On the Formation of Connexions by Compound Eyes in Xenopus. Journal of Physiology 176: 409-417.

Gaze, R. M., and S. C. Sharma. 1970. Axial Differences in the Re-Innervation of the Goldfish Optic Tectum by Regenerating Optic Nerve Fibers. Experimental Brain Research 10: 171-181.

Gazzaniga, M. S. 1992. Nature's Mind: The Biological Roots of Thinking, Emotions, Sexuality, Language, and Intelligence. New York: Basic Books.

Godfrey-Smith, P. 1992. Indication and Adaptation. Synthese 92: 283-312.

_____. 1993. Functions: Consensus Without Unity. Pacific Philosophical Quarterly 74: 196-208.

_____.1994. A Modern History Theory of Functions. Noûs 28: 344-62.

Goffman, E. 1961. Asylums. New York: Anchor.

Goldberg, T. E., and D. R. Weinberger. 1995. A Case Against Subtyping in Schizophrenia. Schizophrenia Research 17: 147-152.

Goode, R., and Griffiths, P. E. 1995. The Misuse of Sober's Selection of/Selection for Distinction. Biology and Philosophy 10: 99-108.

Gould, E., P. N. Tanapat, B. Hastings, and T. J. Shors. 1999a. Neurogenesis in Adulthood: A Possible Role in Learning. Trends in Cognitive Sciences 3: 186-192.

Gould, E. P., A. Beylin, P. Tanapat, A. Reeves, and T. J. Shores. 1999b. Learning Enhances Adult Neurogenesis in the Hippocampal Formation. Nature Neuroscience 2: 260-265.

Gould, S. J., and R. Lewontin. 1979. The Spandrels of San Marco and the Panglossian Paradigm. Proceedings of the Royal Society of London 205: 281-288.

Gould, S. J., and Vrba, E. S. 1982. Exaptation: A Missing Term in the Science of Form. Paleobiology 8: 4—15.

Grace, A. A. 1991. Phasic Versus Tonic Dopamine Release and the Modulation of Dopamine System Reponsivity: A Hypothesis for the Etiology of Schizophrenia. Neuroscience 41: 1-24.

_____. 2000. Gating of Information Flow within the Limbic System and the Pathophysiology of Schizophrenia. Brain Research Reviews 31: 330-341.

Grace, A. A., B. S. Bunney, H. Moore, and C. L. Todd. 1997. Dopamine-cell Depolarization Block as a Model for the Therapeutic Actions of Antipsychotic Drugs. Trends in Neurosciences 20: 31-37.

Griffiths, P. E. 1992. Adaptive Explanation and the Concept of a Vestige. In Trees of Life: Essays in Philosophy of Biology, edited by P. Griffiths, 111-131. Dordrecht: Kluwer.

_____. 1993. Functional Analysis and Proper Function. British Journal for the Philosophy of Science 44: 409-22.

_____. 2005. Function, Homology, and Character Individuation. Philosophy of Science, forthcoming.

Gur, R. C., and R. E. Gur. 1995. Hypofrontality in Schizophrenia: RIP. Lancet 345: 1383-1384.

Hamburger, V. 1958. Regression versus Peripheral Control of Differentiation in Motor Hypoplasia. American Journal of Anatomy 102: 365-410.

_____. 1975. Cell Death in the Development of the Lateral Motor Column of the Chick Embryo. Journal of Comparative Neurology 160: 535-546.

Hamburger, V., and R. Levi-Montalcini. 1949. Proliferation, Differentiation, and Degeneration in the Spinal Ganglia of the Chick Embryo under Normal and Experimental Conditions. Journal of Experimental Zoology 111: 457 – 507.

Hardcastle, V. G. 1999. Understanding Functions: A Pragmatic Approach. In Where Biology Meets Psychology: Philosophical Essays, edited by V. G. Hardcastle, 27-43. Cambridge, MA.: MIT Press.

_____. 2002. On the Normativity of Functions. In Functions: New Essays in the Philosophy of Psychology and Biology, edited by A. Ariew, R. Cummins, and M. Perlman, 144-156. Oxford: Oxford University Press.

Hare, R. M. 1952. The Language of Morals. New York: Oxford University Press.

_____. 1963. Freedom and Reason. Oxford: Clarendon Press.

Harrison, P. J. 1997. Schizophrenia: A Disorder of Neurodevelopment? Current Opinion in Neurobiology 7: 285-289.

_____. 1999. The Neuropathology of Schizophrenia: A Critical Review of the Data and Their Interpretation. Brain 122: 593-624.

Heinrichs, R. W. 2001. In Search of Madness: Schizophrenia and Neuroscience. Oxford: Oxford University Press.

Hempel, C. G. 1965 (1959). The Logic of Functional Analysis. In Aspects of Scientific Explanation, edited by C. G. Hempel, 297-330. New York: Free Press.

Hippocrates. 1952. Hippocratic Writings. Chicago: Encyclopedia Britannica.

Hirsch, S. R., I. Das, L. J. Garey, and J. de Belleroche. 1997. A Pivotal Role for Glutamate in the Pathogenesis of Schizophrenia, and its Cognitive Dysfunction. Pharmacology Biochemistry and Behavior 56: 797-802.

Hoffman, R. E., and T. H. McGlashan. 1993. Parallel Distributed Processing and the Emergence of Schizophrenic Symptoms. Schizophrenia Bulletin 19: 119-140.

Holden, C. 2003. Excited by Glutamate. Science 300: 1866-1868.

Hollyday, M., and V. Hamburger. 1976. Reduction in Naturally Occurring Motor Neuron Loss by Enlargement of the Periphery. Journal of Comparative Neurology 170: 311 – 20.

Horan, B. 1989. Functional Explanations in Sociobiology. Philosophy and Biology 4: 131-158.

Huang, E. J., and L. F. Reichardt. 2001. Neurotrophins: Roles in Neuronal Development and Function. Annual Review of Neuroscience 24: 677-736.

Hubel, D. H., and T. N. Wiesel. 1965. Binocular Interaction in Striate Cortex of Kittens Reared with Artificial Squint. Journal of Neurophysiology 28: 1041-1059.

_____. 1972. Laminar and Columnar Distribution of Geniculo-Cortical Fibers in the Macaque Monkey. Journal of Comparative Neurology 146: 421-450.

Hubel, D. H., T. N. Wiesel, and S. LeVay. 1977. Plasticity of Ocular Dominance Columns in Monkey Striate Cortex. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 278: 377-409.

Hughes, A. 1965. A Quantitative Study of the Development of the Nerves in the Hind-limb of *Eleutherodactylus martinicensis*. Journal of Embryology and Experimental Morphology 9: 269-84.

Hull, D. 1973. Philosophy of Biological Science. Englewood Cliffs, N. J.: Prentice-Hall.

_____. 1988. A Mechanism and its Metaphysics: An Evolutionary Account of the Social and Conceptual Development of Science. Biology and Philosophy 3: 123-155.

Huttenlocher, P. R. 1979. Synaptic Density in the Human Frontal Cortex: Developmental Changes and Effects of Aging. Brain Research 163: 195-205.

Ingvar, D. H. 1987. Evidence for Frontal/Prefrontal Cortical Dysfunction in Chronic Schizophrenia: The Phenomenon of "Hypofrontality" Reconsidered. In Biological Perspectives of Schizophrenia, edited by H. Helmchen, and F. A. Henn, 201-211. Chichester: John Wiley and Sons.

Ingvar, D. H., and G. Franzen. 1974. Abnormalities of Cerebral Blood Flow Distribution in Patients with Chronic Schizophrenia. Acta Psychiatrica Scandinavica 50: 425-462.

Jablensky, A. 2001. Symptoms of Schizophrenia. In Contemporary Psychiatry (Vol. 3), edited by F. Henn, N. Sartorius, H. Helmchen, and H. Lauter, 3-36. Berlin: Springer-Verlag.

Jacobson, M. 1991. Developmental Neurobiology. 3rd ed. New York: Plenum Press.

Jerne, N. K. 1967. Antibodies and Learning: Selection vs. Instruction. In The Neurosciences: A Study Program, edited by G. C. Quarton, T. Melnechuk, and F. O. Schmitt, 200-05. New York: Rockefeller University Press.

Johnson Jr., E. M., and T. L. Deckworth. 1993. Molecular Mechanisms of Developmental Neuronal Death. Annual Review of Neuroscience 16: 31-46.

Jones, E. G. 1995. Cortical Development and Neuropathology in Schizophrenia. In Development of the Cerebral Cortex, edited by G. Bock, and G. Cardew, 277-295. Chichester: John Wiley and Sons.

Kandel, E. R., J. H. Schwartz, and T. M. Jessell. 2000. Principles of Neural Science. 4th ed. New York: McGraw-Hill.

Kapur, S. 2003. Psychosis as a State of Aberrant Salience: A Framework Linking Biology, Phenomenology, and Pharmacology in Schizophrenia. American Journal of Psychiatry 160: 13-23.

Kapur, S., and G. Remington. 2001. Dopamine D_2 Receptors and Their Role in Atypical Antipsychotic Action: Still Necessary and May Even be Sufficient. Biological Psychiatry 50: 873-883.

Kapur, S., and P. Seeman. 2001. Does Fast Dissociation from the Dopamine D_2 Receptor Explain the Action of Atypical Antipsychotics?: A New Hypothesis. American Journal of Psychiatry 158: 360-368.

Katz, L.C., and J. C. Shatz. 1996. Synaptic Activity and the Construction of Cortical Circuits. Science 234: 1133-1138.

Katz, M., J. O. Cole, and H. A. Lowery. 1969. Studies of the Diagnostic Process: The Influence of Symptom Perception, Past Experience, and Ethnic Background on Diagnostic Decisions. American Journal of Psychiatry 125: 937-947.

Kauffman, S. A. 1970. Articulation of Parts Explanation in Biology and the Rational Search for Them. Boston Studies in the Philosophy of Science 8: 257-72.

Kelley, A. E., and K. C. Berridge. 2002. The Neuroscience of Natural Rewards: Relevance to Addictive Drugs. The Journal of Neuroscience 22: 3306-3311.

Kendell, R. E. 1975a. The Role of Diagnosis in Psychiatry. Oxford: Blackwell Scientific Publications.

_____. 1975b. The Concept of Disease and its Implications for Psychiatry. British Journal of Psychiatry 127: 305-15.

_____. 1986. What are Mental Disorders? In Issues in Psychiatric Classification: Science, Practice, and Social Policy, edited by A. M. Freedman, R. Brotman, I. Silverman, and D. Huston, 23-45. New York: Human Sciences Press.

Kendell, R. E., J. E. Cooper, A. J. Gourlay, J. R. M. Copeland, L. Sharpe, and B. J. Gurland. 1971. Diagnostic Criteria of American and British Psychiatrists. Archives of General Psychiatry 25: 123-130.

Kerr, J. F. R., A. H. Wyllie, and A. R. Currie. 1972. Apoptasis: A Basic Biological Phenomenon with Wide-Ranging Implications in Tissue Kinetics. British Journal of Cancer 26: 23-257.

Keshavan, M. S., S. Anderson, and J. W. Pettigrew. 1994. Is Schizophrenia Due to Excessive Synaptic Pruning in the Prefrontal Cortex? A Feinberg Hypothesis Revisited. Journal of Psychiatric Research 28: 239-265.

Kettlewell, B. 1973. The Evolution of Melanism: The Study of a Recurring Necessity. Oxford: Clarendon Press.

Kitcher, P. 1993. Function and Design. Midwest Studies in Philosophy 18: 379-97.

Kirk, S. A., and H. Kutchins. 1992. The Selling of DSM. New York: Aldine de Gruyter.

Kirmayer, L. J., and A. Young. 1999. Culture and Context in the Evolutionary Concept of Mental Disorder. Journal of Abnormal Psychology 108: 446-452.

Klein, D. F. 1978. A Proposed Definition of Mental Illness. In Critical Issues in Psychiatric Diagnosis, edited by R. L. Spitzer, and D. F. Klein, 41-71. New York: Raven Press.

Klerman, G. L., 1978. The Evolution of a Scientific Nosology. In Schizophrenia: Science and Practice, edited by J. C. Shershow, 99-121. Cambridge, MA: Harvard University Press.

Kraepelin, E. 1896. Psychiatrie: Ein Lehrbuch für Studirende und Aerzte. Leipzig: Johann Ambrosius Barth.

_____. 1981 [1907]. Clinical Psychiatry. Delmar, N. Y.: Scholars' Facsimiles and Reprints.

Kripke, S. 1972. Naming and Necessity. Oxford: Blackwell.

Kuhn, T. S. 1977. Objectivity, Value Judgement, and Theory Choice. In The Essential Tension: Selected Studies in Scientific Tradition and Change, edited by T. S. Kuhn, 320-39. Chicago: University of Chicago Press.

Kutchins, H., and S. A. Kirk. 1997. Making Us Crazy: The Psychiatric Bible and the Creation of Mental Disorders. New York: Free Press.

Laing, R. D., and A. Esterson. 1964 Sanity, Madness, and the Family. New York: Basic Books.

LeDoux, J. 2002. Synaptic Self: How our Brains Become Who We Are. New York: Penguin

Lehman, H. 1965. Functional Explanation in Biology. Philosophy of Science 32: 1-20.

Lennox, J. G. 1993. Darwin *was* a Teleologist. Biology and Philosophy 8: 409-21.

Lettvin, J. Y., H. R. Maturana, W. S. McCulloch, and W. H. Pitts. 1959. What the Frog's Eye Tells the Frog's Brain. Proceedings of the IRE 47: 1940-1959.

Levi-Montalcini, R., and S. Cohen. 1960. Effects of the Extract of the Mouse Submaxillary Glands on the Sympathetic System of Mammals. Annals of the New York Academy of Sciences 85: 324-341.

Lewens, T. 2004. Organisms and Artifacts: Design in Nature and Elsewhere. Cambridge, Mass.: MIT Press.

Lewis, D. A. 1997. Development of the Prefrontal Cortex During Adolescence: Insights into Vulnerable Neural Circuits in Schizophrenia. Neuropsychopharmacology 16: 385-398.

Lewontin, R. C. 1970. The Units of Selection. Annual Review of Ecology and Systematics 1: 1-18.

_____. 1998. The Evolution of Cognition: Questions We Will Never Answer. In Invitation to Cognitive Sciences. Vol. 4, Methods, Models, and Conceptual Issues. 2nd ed., edited by D. Scarborough, and S. Sternberg, 107-132.

Lichtman, J. W. 1977. The Reorganization of Synaptic Connexions in the Rat Submandibular Ganglion During Post-Natal Development. Journal of Physiology 273: 155-177.

_____. 1980. On the Predominantly Single Innervation of Submandibular Ganglion Cells in the Rat. Journal of Physiology 302: 121-130.

Lichtman, J. W., S. J. Burden, S. M. Culican, and R. O. L. Wong. 1999. Synapse Formation and Elimination. In Fundamental Neuroscience, edited by M. J. Zigmond, F. E. Bloom, S. C. Landis, J. L. Roberts, and L. R. Squire, 547-580. San Diego: Academic Press.

Liddle, P. F. 1987. The Symptoms of Chronic Schizophrenia: A Re-examination of the Positive-Negative Dichotomy. British Journal of Psychiatry 151: 145-151.

Lilienfeld, S. O., and L. Marino. 1995. Mental Disorder as a Roschian Concept: A Critique of Wakefield's "Harmful Dysfunction" Analysis. Journal of Abnormal Psychology 104: 411-420.

_____. 1999. Essentialism Revisited: Evolutionary Theory and the Concept of Mental Disorder. Journal of Abnormal Psychology 108: 400-411.

Lorenz, K. 1966 (1963). On Aggression. New York: Harcourt, Brace & World.

Mace, C. A. 1949 (1935). Mechanical and Teleological Causation. In Readings in Philosophical Analysis, edited by H. Feigl, and W. Sellars, 534-9. New York: Appleton-Century-Crofts, Inc.

Macklin, R. 1973. The Medical Model in Psychoanalysis and Psychotherapy. Comprehensive Psychiatry 14: 49-69.

Malun, D., and P. C. Brunjes. 1996. Development of Olfactory Glomeruli: Temporal and Spatial Interactions Between Olfactory Receptor Axons and Mitral Cells in Opossums and Rats. Journal of Comparative Neurology 368: 1-16.

Manier, E. 1971. Functionalism and the Negative Feedback Model in Biology. Boston Studies in the Philosophy of Science 8: 225-240.

Marks, I. M., and R. Nesse. 1994. Fear and Fitness: An Evolutionary Analysis. Ethology and Sociobiology 15: 247-261.

Matthews, E. 2003. How Can a Mind be Sick? In Nature and Narrative: An Introduction to the New Philosophy of Psychiatry, edited by K. W. M. Fulford, K. Morris, J. Z. Sadler, and G. Stanghellini, 75-92. Oxford: Oxford University Press.

Mattick, J. S. 2004. The Hidden Genetic Program of Complex Organisms. Scientific American 291 (4) 60-67.

Mayr, E. 1961. Cause and Effect in Biology. Science 134: 1501-06.

McClure, R. K., and D. R. Weinberger. 2001. The Neurodevelopmental Hypothesis of Schizophrenia: A Review of the Evidence. In Current Issues in the Psychopharmacology of Schizophrenia, edited by A. Breier, P. V. Tran, J. M. Herrera, G. D. Tollefson, and F. P. Bymaster, 27-56. Philadelphia: Lippincott Williams and Wilkins.

McGinn, C. 1977. Charity, Interpretation, and Belief. Journal of Philosophy 74: 521-35.

McGlashan, T. H., and R. E. Hoffman. 2000. Schizophrenia as a Disorder of Developmentally Reduced Synaptic Connectivity. Archives of General Psychiatry 57: 637-648.

McLaughlin, P. 2001. What Functions Explain: Functional Explanation and Self-Reproducing Systems. Cambridge: Cambridge University Press.

McNally, R. J. 1994. Panic Disorder: A Critical Analysis. New York: The Guilford Press.

McReynolds, W. T. 1979. DSM-III and the Future of Applied Social Science. Professional Psychology 10: 123-32.

Megone, C. 2000. Mental Illness, Human Function, and Values. Philosophy, Psychiatry, and Psychology 7: 45-65.

Meltzer, H. Y., and A. Y. Deutch. 1999. Neurochemistry of Schizophrenia. In Basic Neurochemistry: Molecular, Cellular, and Medical Aspects, 6th ed., edited by G. J. Siegel, B. W. Agranoff, R. W. Albers, S. K. Fisher, and M. D. Uhler, 1053-1072. Philadelphia: Lippincott-Raven Publishers.

Meyer, C. 1993. Functional Groups of Microorganisms. In Biodiversity and Ecosystem Function, edited by E. Schulze, and H. A. Mooney, 67-96. Berlin: Springer.

Meyer, R. L. 1998. Roger Sperry and his Chemoaffinity Hypothesis. Neuropsychologia 36: 957-980.

Meyer, R. L., and R. W. Sperry. 1976. Retinotectal Specificity: Chemoaffinity Theory. In Studies on the Development of Behavior and the Nervous System. Vol. 3: Neural

and Behavioral Specificity, edited by G. Gottlieb, 111-149. New York: Academic Press.

Millikan, R. G. 1984. Language, Thought, and other Biological Categories. Cambridge, Mass.: MIT Press.

_____. 1989a. An Ambiguity in the Notion 'Function'. Biology and Philosophy 4: 172-76.

_____. 1989b. In Defense of Proper Functions. Philosophy of Science 56: 288-302.

_____. 1993. White Queen Psychology and Other Essays for Alice. Cambridge, MA.: MIT Press.

Millon, T. 1975. Reflections on Rosenhan's "On Being Sane in Insane Places". Journal of Abnormal Psychology 84: 456-61.

_____. 1983. The DSM-III: An Insider's Perspective. American Psychologist 38: 804-15.

_____. 1986. On the Past and Future of the DSM-III: Personal Recollections and Projections. In Contemporary Directions in Psychopathology: Toward the DSM-IV, edited by T. Millon, and G. L. Klerman, 29-70. New York: Guilford Press.

Mills, S. K., and J. H. Beatty. 1979. The Propensity Interpretation of Fitness. Philosophy of Science 46: 263-86.

Mitchell, S. D. 1993. Dispositions or Etiologies? A Comment on Bigelow and Pargetter. Journal of Philosophy 90: 249-59.

_____. 1995. Function, Fitness, and Disposition. Biology and Philosophy 10: 39-54.

Moises, H. W., and I. I. Gottesman. 2001. Genetics, Risk Factors, and Personality Factors. In Contemporary Psychiatry (Vol. 3), edited by F. Henn, N. Sartorius, H. Helmchen, and H. Lauter, 48-59. Berlin: Springer-Verlag.

Moldin, S. O. 1997. The Maddening Hunt for Madness Genes. Nature Genetics 17: 127-129.

Moore, M. S. 1978. Discussion of the Spitzer-Endicott and Klein Proposed Definitions of Mental Disorder (Illness). In Critical Issues in Psychiatric Diagnosis, edited by R. L. Spitzer, and D. F. Klein, 85-104. New York: Raven Press.

Murray, R. M., and S. W. Lewis. 1987. Is Schizophrenia a Neurodevelopmental Disorder? British Medical Journal 295: 681-682.

Murphy, D., and S. Stich. 2000. Darwin in the Madhouse: Evolutionary Psychology and the Classification of Mental Disorders. In Evolution and the Human Mind: Modularity, Language, and Meta-Cognition, edited by P. Carruthers, and A. Chamberlain, 62-92. Cambridge: Cambridge University Press.

Murphy, D., and R. L. Woolfolk. 2001. The Harmful Dysfunction Analysis of Mental Disorder. Philosophy, Psychiatry, and Psychology 7: 241-252.

Murphy, J. M. 1978. The Recognition of Psychosis in Non-Western Societies. In Critical Issues in Psychiatric Diagnosis, edited by R. L. Spitzer, and D. F. Klein, 1-13. New York: Raven Press.

Naeem, S., and S. Li. 1997. Biodiversity Enhances Ecosystem Reliability. Nature 390: 507-509.

Nagel, E. 1953. Teleological Explanation and Teleological Systems. In Vision and Action, edited by S. Ratner, 537-58. New Brunswick, New Jersey: Rutgers University Press.

_____. 1961. The Structure of Science. New York: Harcourt, Brace, & World.

_____. 1977. Teleology Revisited. Journal of Philosophy 76: 261-301.

Neander, K. 1983. Abnormal Psychobiology. Ph.D. diss., La Trobe.

_____. 1991a. Functions as Selected Effects: The Conceptual Analyst's Defense. Philosophy of Science 58: 168-84.

_____. 1991b. The Teleological Notion of 'Function'. Australasian Journal of Philosophy 69: 454-68.

_____. 1995. Misrepresenting and Malfunctioning. Philosophical Studies 79: 109-141.

Nesse, R. M. 1999. Testing Evolutionary Hypotheses about Mental Disorders. In Evolution in Health and Disease, edited by S. C. Stearns, 260-66. Oxford: Oxford University Press.

Nesse, R. M., and G. C. Williams. 1994. Why We Get Sick: The New Science of Darwinian Medicine. New York: Times Books.

_____. 1997. Are Mental Disorders Diseases? In The Maladapted Mind: Classic Readings in Evolutionary Psychopathology, edited by S. Baron-Cohen, 1-22. Hove, UK: Psychology Press.

Nissen, L. 1980-81. Nagel's Self-Regulation Analysis of Teleology. The Philosophical Forum 12: 128-38.

_____. 1997. Teleological Language in the Life Sciences. Lanham, MD.: Rowman and Littlefield.

Oppenheim, R. W. 1981. Neuronal Cell Death and Some Related Regressive Phenomena During Neurogenesis: A Selective Historical Review and Progress Report. In Studies in Developmental Neurobiology: Essays in Honor of Viktor Hamburger, edited by W. M. Cowan, 74-133. Oxford: Oxford University Press.

_____. 1989. The Neurotrophic Theory and Naturally Occurring Motoneuron Death. Trends in Neurosciences 12: 252-255.

_____. 1991. Cell Death During Development of the Nervous System. Annual Review of Neuroscience 14: 453-501.

Oppenheim, R. W., D. Prevette, M. Tyrell, and S. Homma. 1990. Naturally Occurring and Induced Cell Death in the Chick Embryo In Vivo Requires Protein and RNA Synthesis: Evidence for the Role of Cell Death Genes. Developmental Biology 138: 104-113.

Olney, J. W., J. W. Newcomer, and N. B. Farber. 1999. NMDA Receptor Hypofunction Model of Schizophrenia. Journal of Psychiatric Research 33: 523-533.

Papineau, D. 1994. Mental Disorder, Illness and Biological Disfunction. In Philosophy, Psychology and Psychiatry, edited by A. Phillips Griffiths, 73-82. Cambridge: Cambridge University Press.

Parker, G. A. and J. Maynard Smith. 1990. Optimality Theory in Evolutionary Biology. Nature 348: 27-33.

Parsons, T. 1951. The Social System. Glencoe, Ill.: Free Press.

Pâslaru, V. 2005. An Analysis of Functions in Ecosystems. Unpublished ms.

Pettmann, C., and C. E. Henderson. 1998. Neuronal Cell Death. Neuron 20: 653-647.

Pliszka, S. R. 2003. Neuroscience for the Mental Health Clinician. New York: Guilford Press.

Porter, R., and D. Wright, ed. 2003. The Confinement of the Insane: International Perspectives, 1800-1965. Cambridge: Cambridge University Press.

Post, R. M., E. Fink, W. T. Carpenter, and F. K. Goodwin. 1975. Cerebrospinal Fluid Amine Metabolites in Acute Schizophrenia. Archives of General Psychiatry 32: 1063-1069.

Price, J., Sloman, L., Gardner, R., Jr., Gilbert, P., and Rohde, P. 1994. The Social Competition Hypothesis of Depression. British Journal of Psychiatry 164: 309-315.

Prior, E. W. 1985. What is Wrong with Etiological Accounts of Biological Function? Pacific Philosophical Quarterly 66: 310-28.

Prior, E., R. Pargetter, and F. Jackson. 1982. Three Theses about Dispositions. American Philosophical Quarterly 19: 251-57.

Purves, D. 1977. The Formation and Maintenance of Synaptic Connections. In Function and Formation of Neural Systems, edited by G. S. Stent, 21-49. Berlin: Abakon-Verlagsgesellschaft.

_____. 1988. A New Theory of Brain Function. The Quarterly Review of Biology 63: 202-204.

_____. 1994. Neural Activity and the Growth of the Brain. Cambridge: Cambridge University Press.

Purves, D., and J. W. Lichtman. 1978. Formation and Maintenance of Synaptic Connections in Autonomic Ganglia. Physiological Reviews 58: 821-862.

_____. 1980. Elimination of Synapses in the Developing Nervous System. Science 210: 153-157.

Purves, D., L. E. White, and D. R. Riddle. 1996. Is Neural Development Darwinian? Trends in Neuroscience 19: 460-464.

Purves, D., G. J., Augustine, D. Fitzpatrick, W. C. Hall, A. LaMantia, J. O. McNamara, and S. M. Willams, eds. 2004. Neuroscience, 3rd ed. Sunderland, Mass.: Sinauer.

Putnam, H. 1975. The Meaning of 'Meaning'. In Philosophical Papers, Vol. II: Mind, Language, and Reality, edited by H. Putnam, 215-71. Cambridge: Cambridge University Press.

Quartz, S. R., and T. J. Sejnowski. 1997. The Neural Basis of Cognitive Development: A Constructivist Manifesto. Behavioral and Brain Sciences 20: 537-596.

Rajan, I., and H. T. Cline. 1998. Glutamate Receptor Activity is Required for Normal Development of Tectal Cell Dendrites *In Vivo*. Journal of Neuroscience 18: 7836-7846.

Rakic, P. 1976. Prenatal Genesis of Connections Subserving Ocular Dominance in the Rhesus Monkey. Nature 261: 467-471.

Rakic, P., J. Bourgeois, M. F. Eckenhoff, N. Zecevic, and P. S. Goldman-Rakic. 1986. Concurrent Overproduction of Synapses in Diverse Regions of the Primate Cerebral Cortex. Science 232: 232-235.

Raz, S., and N. Raz. 1990. Structural Brain Abnormalities in the Major Psychoses: A Quantitative Review of the Evidence from Computerized Imaging. Psychological Bulletin 108: 93-108.

Redlich, F. C., and Freedman, D. X. 1966. The Theory and Practice of Psychiatry. New York: Basic Books.

Reeve, H. K. and P. W. Sherman. 1993. Adaptation and the Goals of Evolutionary Research. Quarterly Review of Biology 68: 1-32.

Rich, M. M., H. Colman, and J. W. Lichtman. 1994. In Vivo Imaging Shows Loss of Synaptic Sites from Neuromuscular Junction in a Model of Myasthenia Gravis. Neurology 44: 2138-2145.

Richters, J. E., and D. Cicchetti. 1993. Mark Twain meets DSM-III-R: Conduct Disorder, Development, and the Concept of Harmful Dysfunction. Development and Psychopathology 5: 5-29.

Richters, J. E., and Hinshaw, S. P. 1999. The Abduction of Disorder in Psychiatry. Journal of Abnormal Psychology 108: 438-445

Riley, B. P., and P. McGuffin. 2000. Linkage and Associated Studies of Schizophrenia. American Journal of Medical Genetics (Seminars in Medical Genetics) 97: 23-44.

Robinson, D. N. 2003. Psychiatry and Law. In Nature and Narrative: An Introduction to the New Philosophy of Psychiatry, edited by K. W. M. Fulford, K. Morris, J. Z. Sadler, and G. Stanghellini, 93-101. Oxford: Oxford University Press.

Robinson, K. 1998. Women and Ownership of PMS: The Structuring of a Psychiatric Disorder – Review. The Australian Journal of Anthropology 9: 338-340.

Rosenblueth, A., Wiener, N. and Bigelow, J. 1943. Behavior, Purpose and Teleology. Philosophy of Science 10: 18-24.

Rosenfield, I. 1986. Neural Darwinism: A New Approach to Memory and Perception. The New York Review of Books 9 October, 21-27.

Rosenhan, D. L. 1973. On Being Sane in Insane Places. Science 179: 250-58.

_____. 1975. The Contextual Nature of Psychiatric Diagnosis. Journal of Abnormal Psychiatry 84: 462-474.

Ruse, M. 1971. Functional Statements in Biology. Philosophy of Science 38: 87-95.

_____. 1973. The Philosophy of Biology. London: Hutchinson University Library.

Ryle, G. 1949. The Concept of Mind. London: Hutchinson.

Sadler, J. Z. 1997. Recognizing Values: A Descriptive-Causal Method for Medical/Scientific Discourses. The Journal of Medicine and Philosophy 22: 541-565.

_____. 1999. Horsefeathers: A Commentary on "Evolutionary Versus Prototype Analyses of the Concept of Disorder". Journal of Abnormal Psychology 108: 433-37.

_____. 2004. Values and Psychiatric Diagonsis. Oxford: Oxford University Press.

Sadler, J. Z., and G. J. Agich. 1995. Diseases, Functions, Values, and Psychiatric Classification. Philosophy, Psychiatry, and Psychology 2: 219-231.

Salkovskis, P. M. 1991. The Importance of Behaviour in the Maintenance of Anxiety and Panic: A Cognitive Account. Behavioural Psychotherapy 19: 6-19.

Sarbin, T. R. 1967. On the Futility of the Proposition that Some People be Labeled "Mentally Ill". Journal of Consulting Psychology 31: 447-53.

_____. 1969. The Scientific Status of the Mental Illness Metaphor. In Changing Perspectives in Mental Illness, edited by S. C. Plog, and R. B. Edgerton, 9-31. New York: Holt, Rinehart and Winston.

Sarkar, S. 1996. Ecological Theory and Anuran Declines. BioScience 46: 199-207.

_____. 2005. Molecular Models of Life: Philosophical Papers on Molecular Biology. Cambridge, Mass.: MIT Press.

Sartorius, N., A. Jablensky, A. Korten, G. Ernberg, M. Anker, J. E. Cooper, and R. Day. 1986. Early Manifestations and First-Contact Incidence of Schizophrenia in Different Cultures. Psychological Medicine 16: 909-928.

Schacht, T., and P. E. Nathan. 1977. But is it Good for the Psychologists? Appraisal and Status of DSM III. American Psychologist 32, 1017-1025.

Schaffner, K. F. 1993. Discovery and Explanation in Biology and Medicine. Chicago: University of Chicago Press.

Scheff, T. J. 1966. Being Mentally Ill. Chicago: Aldine.

Scheffler, I. 1966 (1958). Thoughts on Teleology. British Journal for the Philosophy of Science 9: 265-284.

Schlosser, G. 1998. Self-Reproduction and Functionality: A Systems-Theoretical Approach to Teleological Explanation. Synthese 116: 303-354.

Scholes, J. 1979. Nerve Fiber Topography in the Retinal Projection to the Tectum. Nature 278: 620-624.

Schwartz, P. H. 1999. Proper Function and Recent Selection. Philosophy of Science 66 (Proceedings): S210-S222.

_____. 2004. An Alternative to Conceptual Analysis in the Function Debate. The Monist 87: 136-153.

Sedgwick, P. 1981. Illness – Mental and Otherwise. In Concepts of Health and Disease: Interdicsiplinary Perspectives, edited by A. L. Caplan, H. T. Engelhardt, Jr., and J. J. McCartney, 119-129. London: Addison-Wesley.

Shapiro, L. A. 1992. Darwin and Disjunction: Foraging Theory and Univocal Assignments of Content. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1: 469-480.

Sherrington, C. S. 1906. The Integrative Action of the Nervous System. New Haven: Yale University Press.

Shorter, E. 1997. A History of Psychiatry: From the Era of the Asylum to the Age of Prozac. New York: John Wiley and Sons.

Siegler, M., and Osmond, H. 1974. Models of Madness, Models of Medicine. New York: Macmillan.

Silverman, J. 1967. Shamans and Acute Schizophrenia. American Anthropologist 69: 21-31.

Skinner, B. F. 1953. Science and Human Behavior. New York: The Free Press.

_____. 1981. Selection by Consequences. Science 213: 501-504.

Simon, H. A. 1969. The Sciences of the Artificial. Cambridge, MA.: MIT Press.

Sin, W. C., K. Haas, E. S. Ruthazer, and H. T. Cline. 2002. Dendrite Growth Increased by Visual Activity Requires NMDA Receptor and Rho GTPases. Nature 419: 475-480.

Smith, A. C. 1982. Schizophrenia and Madness. London: George Allen and Umwin.

Sommerhoff, G. 1950. Analytical Biology. London: Oxford University Press.

_____. 1969. The Abstract Characteristics of Living Systems. In Systems Thinking, edited by F. E. Emery, 147-202. Middlesex, England: Penguin.

Sorabji, R. 1964. Function. Philosophical Quarterly 14: 289-302.

Sperry, R. W. 1944. Optic Nerve Regeneration with Return of Vision in Anurans. Journal of Neurophysiology 7: 57-69.

_____. 1951. Mechanisms of Neural Maturation. In Handbook of Experimental Psychology, edited by S. S. Stevens, 236-280. New York: Wiley.

_____. 1963. Chemoaffinity in the Orderly Growth of Nerve Fiber Patterns of Connections. Proceedings of the National Academy of Sciences of the United States of America 50: 703-710.

Spitzer, R. L. 1975. On Psuedoscience in Science, Logic in Remission, and Psychiatric Diagnosis: A Critique of Rosenhan's "On Being Sane in Insane Places". Journal of Abnormal Psychology 84: 442-452.

_____.1981. The Diagnostic Status of Homosexuality in *DSM-III*: A Reformulation of the Issues. American Journal of Psychiatry 138: 210-215.

Spitzer, R. L., and J. Endicott. 1978. Medical and Mental Disorder: Proposed Definition and Criteria. In Critical Issues in Psychiatric Diagnosis, edited by R. L. Spitzer, and D. F. Klein, 15-39. New York: Raven Press.

Spitzer, R. L., and J. B. W. Williams. 1982. The Definition and Diagnosis of Mental Disorder. In Deviance and Mental Illness, edited by W. R. Gove, 15-31. Beverly Hills: Sage Publications.

Spitzer, R. L., M. Sheehy, and J. Endicott. 1977. DSM-III: Guiding Principles. In Psychiatric Diagnosis, edited by V. M. Rakoff, H. C. Stancer, and H. B. Kedward, 1-24. New York: Brunner/Mazel.

Spitzer, R. L., and J. Wakefield. 1999. DSM_IV Diagnostic Criterion for Clinical Significance: Does it Help Solve the False Positives Problem? American Journal of Psychiatry 156: 1856-1864.

Sporns, O. 1997a. Variation and Selection in Neural Function. Trends in Neurosciences 20: 291.

_____. 1997b. Deconstructing Neural Constructivism. Behavioral and Brain Sciences 20: 576-577.

Sterelny, K. 1990. The Representational Theory of Mind. Oxford: Blackwell.

Stevens, A., and J. Price. 2000. Evolutionary Psychiatry: A New Beginning. London: Routledge.

Stevenson, C. L. 1937. The Emotive Meaning of Ethical Terms. Mind. 46: 14-31.

Stoller, R., J. Marmor, I. Beiber, et al. 1973. A Symposium: Should Homosexuality be in the APA Nomenclature? American Journal of Psychiatry 130: 1207-16.

Strauss, J. S., W. T. Carpenter, and J. J. Bartko. 1974. The Diagnosis and Understanding of Schizophrenia: III. Speculations on the Processes that Underlie Schizophrenic Symptoms and Signs. Schizophrenia Bulletin 11: 61-69.

Swazey, J. P. 1974. Chlorpromazine in Psychiatry. Cambridge, MA: MIT Press.

Szasz, T. S. 1961. The Myth of Mental Illness: Foundations of a Theory of Personal Conduct. New York: Harper & Row.

Takamori, M. 2004. Lambert-Eaton Myasthenic Syndrome as an Autoimmune Calcium Channelopathy. Biochemical and Biophysical Research Communications 322: 1347-1351.

Thagard, P. 1988. Computational Philosophy of Science. Cambridge, Mass.: MIT Press.

Thorndike, E. L. 1911. Animal Intelligence: Experimental Studies. New York: Macmillan.

Tinbergen, N. 1963. On the Aims and Methods of Ethology. Zeitschrift für Tierpsychologie 20: 410-29.

Torrey, E. F., and R. H. Yolken. 2000. Familial and Genetic Mechanisms in Schizophrenia. Brain Research Reviews 31: 113-117.

Toulmin, S. 1975. Concepts of Function and Mechanism in Medicine and Medical Science. In Evaluation and Explanation in the Biomedical Sciences, edited by H. T. Engelhardt , and S. F. Spicker, 51-66. Dordrecht: D. Reidel.

Tsuang, M. 2000. Schizophrenia: Genes and Environment. Biological Psychiatry 47: 210-220.

Tsuang, M. T., and S. V. Faraone. 1995. The Case For Heterogeneity in the Etiology of Schizophrenia. Schizophrenia Research 17: 161-175.

Ullman, L. P. and L. Krassner. 1966. Case Studies in Behavior Modification. New York: Holt, Rinehart, and Winston.

Van Kammen, D. P., W. B. van Kammen, L. S. Mann, T. Seppala, and M. Linnoila. 1986. Dopamine Metabolism in the Cerebrospinal Fluid of Drug-Free Schizophrenic Patients With and Without Cortical Atrophy. Archives of General Psychiatry 43: 978-983.

Wakefield, J. C. 1992a. The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values. American Psychologist 47: 373-88.

_____. 1992b. Disorder as Harmful Dysfunction: A Conceptual Critique of DSM-III-R's Definition of Mental Disorder. Psychological Review 99: 232-247.

_____. 1993. Limits of Operationalization: A Critique of Spitzer and Endicott's (1978) Proposed Operational Criteria for Mental Disorder. Journal of Abnormal Psychology 102: 160-72.

_____. 1999a. Evolutionary Versus Prototype Analyses of the Concept of Disorder. Journal of Abnormal Psychology 108: 374-399.

_____. 1999b. Mental Disorder as a Black Box Essentialist Concept. Journal of Abnormal Psychology 108: 465-472.

Wakefield, J. C. and M. B. First. 2003. Clarifying the Distinction Between Disorder and Nondisorder: Confronting the Overdiagnosis (False-Positives) Problem in DSM-V. In Advancing DSM: Dilemmas in Psychiatric Diagnosis, edited by K. A. Phillips, M. B. First, and H. A. Pincus, 23-55. Washington D.C.: American Psychiatric Association.

Walicke, P. A. 1989. Novel Neurotrophic Factors, Receptors, and Oncogenes. Annual Review of Neuroscience 12: 103-126.

Walker, E., and R. J. Lewine. 1988. The Positive/Negative Symptom Distinction in Schizophrenia. Validity and Etiological Relevance. Schizophrenia Research 1: 315- 328.

Walsh, D. M. 1996. Fitness and Function. British Journal for the Philosophy of Science 47: 553-74.

Walsh, D. M., and A. Ariew. 1996. A Taxonomy of Functions. Canadian Journal of Philosophy 26: 493-514.

Wang, G. Y., L. C. Liets, and L. M. Chalupa. 2001. Unique Functional Properties of On and Off Pathways in the Developing Mammalian Retina. Journal of Neuroscience 21: 4310-4317.

Weinberger, D. R. 1984. Computed Tomography (CT) Findings in Schizophrenia: Speculation on the Meaning of it All. Journal of Psychiatric Research 18: 477-490.

_____. 1987. Implications of Normal Brain Development for the Pathogenesis of Schizophrenia. Archives of General Psychiatry 44: 660-669.

Weinberger, D. R., K. F. Berman, and R. F. Zec. 1986. Physiologic Dysfunction of Dorsolateral Prefrontal Cortex in Schizophrenia: I. Regional Cerebral Blood Flow Evidence. Archives of General Psychiatry 43: 114-124.

Whitaker, R. 2002. Mad in America. Cambridge, MA: Perseus.

Widiger, T. A., and T. J. Trull. 1991. Diagnosis and Clinical Assessment. Annual Review of Psychology 42: 109-33.

Wiesel, T. N., and D. H. Hubel. 1963. Single-Cell Responses in Striate Cortex of Kittens Deprived of Vision in One Eye. Journal of Neurophysiology 26: 1003-1017.

Wilcox, R.E., R. A. Gonzales, and J. D. Miller. 1999. Introduction to Neurotransmitters, Receptors, Signal Transduction, and Second Messengers. In Textbook of Pharmacology, edited by A. F. Schatzberg, and C. B. Nemeroff, 3-36. Washington D.C.: American Psychiatric Press.

Williams, G. C. 1966. Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought. Princeton: Princeton University Press.

Wilson, M. 1993. DSM-III and the Transformation of American Psychiatry: A History. American Journal of Psychiatry 150: 399-410.

Wimsatt, W. C. 1971. Some Problems with the Concept of Feedback. Boston Studies in the Philosophy of Science 8: 241-256.

_____. 1972. Teleology and the Logical Structure of Function Statements. Studies in History and Philosophy of Science 3: 1-80.

_____. 2002. Functional Organization, Analogy, and Inference. In Functions: New Essays in the Philosophy of Psychology and Biology, edited by A. Ariew, R. Cummins, and M. Perlman, 173-221. Oxford: Oxford University Press.

Wise, R. A. 2002. Brain Reward Circuitry: Insights from Unsensed Incentives. Neuron 36: 229-240.

Wittchen, H. A., and C. A. Essau. 1991. The Epidemiology of Panic Attacks, Panic Disorder and Agoraphobia. In Panic Disorder and Agoraphobia, edited by J. R. Walker, G. R. Norton, and C. A. Ross. Pacific Grove, CA.: Brooks/Cole.

Wong, R. O. L., and A. Ghosh. 2002. Activity-Dependent Regulation of Dendritic Growth and Patterning. Nature Reviews Neuroscience 3: 803-812.

Woodfield, A. 1976. Teleology. Cambridge: Cambridge University Press.

Woodruff, R. A., D. W. Goodwin, and S. B. Guze. 1974. Psychiatric diagnosis. New York: Oxford University Press.

Woods, B. T., D. Yurgelun-Todd, J. M. Goldstein, L. J. Seidman, and M. T. Tsuang. 1996. MRI Brain Abnormalities in Chronic Schizophrenia: One Process or More? Biological Psychiatry 40: 585-596.

Woolfolk, R. L. 1999. Malfunction and Mental Disorder. The Monist 82: 658-670.

World Health Organization. 1992. The ICD-10 Classification of Mental and Behavioral Disorders. Geneva: World Health Organization.

Wouters, A. 2003. Four Notions of Biological Function. Studies in History and Philosophy of Biological and Biomedical Sciences 34: 633-668.

_____. 2005a. The Functional Perspective in Organismic Biology. In Current Themes in Theoretical Biology, edited by T. A. C. Reydon, and L. Hemerik, 33-69. Dordrecht: Springer.

_____. 2005b. The Function Debate in Philosophy. Acta Biotheoretica 53: 123-151.

Wright, L. 1972. A Comment on Ruse's Analysis of Function Statements. Philosophy of Science 39: 512-14.

_____. 1973. Functions. Philosophical Review 82: 139-168.

_____. 1976. Teleological Explanations: An Etiological Analysis of Goals and Functions. Berkeley: University of California Press.

Yuan, J., and H. R. Horvitz. 1990. The *Caenorhabditis elegans* Genes *ced-3* and *ced-4* Act Cell Autonomously to Cause Programmed Cell Death. Developmental Biology 138: 33-41.

Zubin, J. 1977. But is it Good for Science? The Clinical Psychologist 31:1-7.

Zuckermann, M. 1999. Vulnerability to Psychopathology, Washington D.C.: American Psychological Association.

Vita

Justin Garson was born on May 22, 1973 in Washington D.C., the son of John Richard Garson and Marjorie Ann Leary. He received his Bachelor of Arts from the Evergreen State College in Olympia, Washington in the Spring of 1998 and his Masters of Arts in philosophy from the University of Texas at Austin in the Fall of 2002. While writing his dissertation he received a Liberal Arts Continuing Fellowship from the University of Texas at Austin, Fall, 2003 – Spring, 2004, and a Liberal Arts Graduate Research Fellowship from the Graduate School at the University of Texas at Austin, Summer, 2003. He was an assistant editor for *The Philosophy of Science: An Encyclopedia* (Routledge, 2006), and held several teaching assistantships. He also held a research internship at the Biodiversity and Biocultural Conservation Laboratory at the University of Texas at Austin, and has authored or co-authored several articles in the history and philosophy of science, and conservation biology.

Permanent address: 301 West 38th St. Apt. 206, Austin, Texas, 70705

This dissertation was typed by the author.